

SOEPpapers

on Multidisciplinary Panel Data Research

371

Robert Maderitsch

**Eine ökonometrische Analyse
der Liquiditätsbeschränkung deutscher Haushalte
im Lichte der US-Immobilienkrise**

Berlin, April 2011

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
<http://www.diw.de/soeppapers>

Editors:

Georg **Meran** (Dean DIW Graduate Center)

Gert G. **Wagner** (Social Sciences)

Joachim R. **Frick** (Empirical Economics)

Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics)

Christoph **Breuer** (Sport Science, DIW Research Professor)

Elke **Holst** (Gender Studies)

Martin **Kroh** (Political Science and Survey Methodology)

Frieder R. **Lang** (Psychology, DIW Research Professor)

Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)

C. Katharina **Spieß** (Educational Science)

Martin **Spieß** (Survey Methodology, DIW Research Professor)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Eine ökonometrische Analyse der Liquiditätsbeschränkung deutscher Haushalte im Lichte der US-Immobilienkrise

Robert Maderitsch¹

Zusammenfassung

In dieser Arbeit kommt ein neuer ökonometrischer Ansatz für Paneldaten zum Einsatz, welcher den Umgang mit persistenter, unbeobachtbarer Heterogenität in allgemeinen nichtseparablen und nonparametrischen Konstellationen ermöglicht. Der Ansatz geht zurück auf Hoderlein und White (2009). Diese untersuchen für die USA, welchen Einfluss marginale Einkommenszuwächse auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, haben. Für die Jahre kurz vor der Immobilienkrise zeigen sie, dass der Erwerb von Wohneigentum vom laufenden Einkommen unabhängig wird. Die Rede ist von einer verschwindenden Liquiditätsbeschränkung.

In dieser Arbeit wird mit Hilfe der SOEP-Daten untersucht ob der gleiche Zusammenhang auch für deutsche Haushalte festgestellt werden kann. Es zeigt sich, dass der durchschnittliche marginale Effekt des Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, durchweg positiv und über die Zeit hinweg nahezu konstant ist. Im Gegensatz zu den USA, ist der Effekt jedoch nicht über alle Einkommensbereiche hinweg, sondern nur im Bereich der mittleren Einkommen festzustellen. Besonders ausgeprägt scheint er außerdem für jüngere Haushalte zu sein. Für ein Verschwinden oder eine Reduktion der Liquiditätsbeschränkung findet sich insgesamt keine Evidenz.

JEL Klassifikation: C14, C23, D10, E20

Schlüsselwörter: Wohneigentum, Haushalte, Nonparametrisch, Panel Daten, Fixed Effects

Autor:

Robert Maderitsch

Universität Erfurt

Nordhäuser Straße 63

D-99089 Erfurt

robert.maderitsch@uni-erfurt.de

+493617374592

¹Die folgende Arbeit entspricht meiner am 21.12.2010 abgegebenen Diplomarbeit. Mein herzlicher Dank für die hervorragende Betreuung geht an dieser Stelle an Eva Schlenker, Prof. Dr. Gerhard Wagenhals und Prof. Ph.D. Stefan Hoderlein.

Inhaltsverzeichnis

Abbildungsverzeichnis	II
Tabellenverzeichnis	III
Abkürzungsverzeichnis	III
1. Einführung	1
1.1. Motivation	1
1.2. Aufbau der Arbeit	3
2. Theorieteil	5
2.1. Binäre Regression	5
2.2. Paneldaten und Fixed Effects	8
2.3. Semi- und nonparametrische Verfahren	13
2.3.1. Überblick	13
2.3.2. Kernel-Regression	14
2.3.3. Lokale polynomiale Regression	21
2.4. Beitrag von Hoderlein und White (2009)	22
2.4.1. Die Idee	22
2.4.2. Angrenzende Literatur	24
2.4.3. Identifikation marginaler Effekte in nichtseparablen Funktionen	25
2.4.4. Endogenes Modell binärer Wahl mit Heterogenität	37
3. Empirieteil	43
3.1. Immobilienkrise und Liquiditätsbeschränkung der Haushalte	43
3.1.1. Wirtschafts- und Finanzkrise von 2007	43
3.1.2. Bisherige Risikomaße für den Immobilienbereich	44
3.1.3. Rolle der Haushalte	46
3.1.4. Beitrag dieser Arbeit	48
3.2. Betrachtung von Hoderlein und White (2009) für die USA	51
3.2.1. Ökonometrische Umsetzung	51
3.2.2. Resultate	53
3.3. Vergleichende Untersuchung für Deutschland	57
3.3.1. Datengrundlage	57
3.3.2. Datenaufbereitung	59
3.3.3. Schätzung	63
3.3.4. Ergebnisse	64
4. Fazit	71
Literaturverzeichnis	IV
A. Anhang	X
A.1. Deskriptive Statistiken für die Originaldaten	X
A.2. Deskriptive Statistiken für die aufbereiteten Daten	XII
A.3. Resultate für die USA	XIV

A.4. Übersicht Ergebnisse Deutschland	XV
A.5. Ergebnisse für Deutschland 1999/2001 im Detail	XVII

Abbildungsverzeichnis

1. Preisentwicklung der US-Immobilien (Case-Shiller-Index, Januar . . .	45
2. Marginale Wahrscheinlichkeit für den Besitz von Wohneigentum bei gegebenem Einkommen (Zeitraum: 1999/2001).	53
3. Marginale Wahrscheinlichkeit für den Besitz von Wohneigentum bei gegebenem Einkommen (Zeitraum: 2003/2005).	55
4. Verteilung der Haushalts-Nettoeinkommen 1999.	60
5. Verteilung der Variable Alter.	61
6. Ergebnisse für die Gruppe „um Durchschnitt“ 1999/2001.	66
7. Ergebnisse für die Gruppen „jung“ und „alt“ 1999/2001.	67
8. Ergebnisse für die Gruppe „alle“ 1999/2001.	68
9. Ergebnisse für die Gruppe „alle“ 1999/2001 im Vergleich zu 2003/2005.	69
A1. Verteilung Alter 99.	X
A2. Verteilung der Haushalts-Nettoeinkommen 1999.	X
A3. Verteilung der Haushalts-Nettoeinkommen 2001.	X
A4. Verteilung Alter 2003.	XI
A5. Verteilung der Haushalts-Nettoeinkommen 2003.	XI
A6. Verteilung der Haushalts-Nettoeinkommen 2005.	XI
A7. Alter und Haushalts-Nettoeinkommen 1999.	XII
A8. Logarithmierte Haushalts-Nettoeinkommen 1999.	XII
A9. Verteilung der Haushalts-Nettoeinkommen 2001.	XII
A10. Alter und Haushalts-Nettoeinkommen 2003.	XIII
A11. Logarithmierte Haushalts-Nettoeinkommen 2003.	XIII
A12. Haushalts-Nettoeinkommen 2005	XIII
A13. Marginale Wahrsch. für den Besitz von Wohneigentum	XIV
A14. Ergebnisse 99/01, alle Altersgruppen.	XVII
A15. Ergebnisse 99/01, Altersgruppe „jung“.	XVIII
A16. Ergebnisse 99/01, Altersgruppe „alt“.	XVIII
A17. Ergebnisse 99/01, Altersgruppe „um Durchschnitt“.	XIX
A18. Ergebnisse 03/05, alle Altersgruppen.	XX
A19. Ergebnisse 03/05, Altersgruppe „jung“.	XXI
A20. Ergebnisse 03/05, Altersgruppe „alt“.	XXI
A21. Ergebnisse 03/05, Altersgruppe „um Durchschnitt“.	XXII

Tabellenverzeichnis

1.	Stichproben im SOEP	58
2.	Die wesentlichen Variablen	59
3.	Änderung des Eigentumsstatus zwischen 1999 und 2001	62
4.	Änderung des Eigentumsstatus zwischen 2003 und 2005	62
5.	Untersuchte Altersgruppen	64
6.	Untersuchte Einkommensänderungen	64
A1.	Ergebnisübersicht Panel 1999/2001	XV
A2.	Ergebnisübersicht Panel 2003/2005	XVI

Abkürzungsverzeichnis

CNEF	Cross National Equivalent File
FD	First Differencing
FE	Fixed Effects
iid	independent, identically distributed bzw. unabhängig und identisch verteilt
LASD	Local Average Structural Derivative bzw. Lokale Durchschnittliche Strukturelle Ableitung
ML	Maximum-Likelihood
PSID	Panel Study of Income Dynamics
SOEP	Sozio-ökonomisches Panel

1. Einführung

1.1. Motivation

„We’re creating...an ownership society in this country, where more Americans than ever will be able to open up their door where they live and say, welcome to my house, welcome to my piece of property.“¹

Aktuell, so schätzt die Wirtschaftswoche, liegt die Zahl der in den USA seit Beginn der Immobilienkrise zwangsversteigerten Wohnimmobilien bei rund 6,2 Millionen. Weitere drei Millionen Zwangsversteigerungen werden bis zum Ende des Jahres 2012 erwartet.²

Als Hauptfaktoren für das Entstehen der Immobilienkrise in den USA gilt die unvorsichtige Kreditvergabe der Immobilienfinanzierer und die im obigen Zitat angesprochene Politik der amerikanischen Regierung der gezielten Förderung von Wohneigentum.

Einen tieferen Einblick in die Ursachen der Immobilienkrise liefert ein aktueller Forschungsbeitrag von Hoderlein und White (2009). Dieser beleuchtet den Zusammenhang zwischen laufendem Haushaltseinkommen und der Wahrscheinlichkeit dafür, dass ein Haushalt über Wohneigentum verfügt. Einige Jahre vor der Immobilienkrise zeigt sich, dass marginale Einkommenszuwächse die Wahrscheinlichkeit, Wohneigentum zu besitzen, signifikant erhöhen. Kurz vor der Immobilienkrise ist dies nicht mehr der Fall. Daraus lässt sich schließen, dass der Erwerb von Wohneigentum vom laufenden Haushaltseinkommen unabhängig geworden ist. Hoderlein und White (2009) sprechen daher von einer verschwindenden Liquiditätsbeschränkung für die amerikanischen Haushalte.

In dieser Arbeit erfolgt eine analoge Untersuchung für deutsche Haushalte. Die entscheidende Frage ist, ob sich ähnliche Resultate wie in den USA ergeben oder nicht. Da in Deutschland bisher keine Krise im Bereich der Wohnimmobilien aufgetreten ist, kann hiervon a priori nicht ausgegangen werden. Dennoch stellt sich die Frage, ob beispielsweise anstatt einer verschwindenden Liquiditätsbeschränkung, eine sich lediglich verringernde Liquiditätsbeschränkung gefunden werden kann.

¹ So waren die Worte des ehemaligen Präsidenten G. W. Bush bei seiner Antrittsrede im Jahr 2004.

² Vgl. Wirtschaftswoche, 25.10.2010, S.56.

Zur Verwendung kommt bei den deutschen Daten der gleiche Ansatz wie bei Hoderlein und White (2009). Aus ökonometrischer Sicht ist an diesem besonders reizvoll, dass die Idee der Fixed Effects auf nichtlineare Zusammenhänge mit unendlich dimensional, unbeobachtbaren Variablen erweitert wird. Somit können beispielsweise individuenspezifische Präferenzen berücksichtigt werden, obwohl für diese keine Daten vorliegen.

Hinzu kommt, dass sowohl ein einfaches Paneldatenmodell, als auch ein nonparametrisches Regressionsverfahren zum Einsatz kommen. Beide Ansätze haben in den letzten Jahren aufgrund der rasanten Entwicklung von Computertechnologie und Rechenkapazität enorm an Bedeutung gewonnen.

Paneldatenmodelle zeichnen sich dadurch aus, dass für die betrachteten Einheiten Informationen über mehrere Perioden (im Längsschnitt) vorliegen. Hierdurch ergeben sich einerseits Effizienzgewinne für die ökonometrischen Schätzungen, andererseits jedoch, wie bei Hoderlein und White (2009), auch ganz neue Identifikationsmöglichkeiten. Aus gegebenen Daten können somit noch mehr Schlüsse gezogen werden.

Die semi- und nonparametrischen Regressionsverfahren haben den Vorteil, dass nur wenige einschränkende Annahmen getroffen werden müssen. Die Wahrscheinlichkeitsverteilung für die abhängige Variable und der funktionale Zusammenhang zwischen abhängiger Variable und unabhängigen Variablen, können prinzipiell im Vorfeld offen gelassen werden. Dies bietet vor allem dann Vorteile, wenn so wie in dieser Arbeit, kein Anlass dazu besteht, lineare Beziehungen zwischen den zu untersuchenden Variablen anzunehmen. Im Vergleich zu parametrischen Verfahren gelingt eine oft bessere Anpassung an die komplexen Datenstrukturen. Die komplizierte Interpretation der Ergebnisse ist insoweit kein Problem, da die Ergebnisse meist grafisch veranschaulicht werden können.

1.2. Aufbau der Arbeit

Grundsätzlich gliedert sich die vorliegende Arbeit in zwei Teile: Der Theorieteil (Abschnitt 2) widmet sich dem relevanten ökonometrischen Hintergrund. Der Empirieteil (Abschnitt 3) beschäftigt sich mit der Anwendung der behandelten Methoden.

Zur Erleichterung des Zugangs zum Beitrag von Hoderlein und White (2009), werden im Theorieteil vorab einige Grundlagen erläutert. Vom fortgeschrittenen Leser können diese übersprungen werden.

Da Hoderlein und White (2009) den Spezialfall des Modells binärer Wahl behandeln, wird in Abschnitt 2.1 zunächst die binäre Regression dargestellt.

Daraufhin wird in Abschnitt 2.2 die Panelstruktur eingeführt, bevor allgemeine Methoden zum Umgang mit zeitkonstanten, unbeobachtbaren Faktoren vorgestellt werden. Diese stellen die Differenzenbildung und die Fixed-Effects-Schätzung dar. Letztere wird für den Fall des binären Modells genauer betrachtet.

In Abschnitt 2.3 werden dann die Grundlagen zur nonparametrischen bzw. semiparametrischen Regression präsentiert. Dabei erfolgt ein allgemeiner Überblick über die gängigen Verfahren (2.3.1). Die wesentlichen Vor- und Nachteile der nonparametrischen Verfahren gegenüber den normalen parametrischen Regressionsverfahren werden herausgearbeitet. Im Detail wird dann auf die in der Arbeit von Hoderlein und White (2009) relevanten Methoden eingegangen: Die Kernel-Regression (2.3.2) und die lokale polynomiale Regression (2.3.3).

Nach dieser Vorarbeit wird in Abschnitt 2.4 der Beitrag von Hoderlein und White (2009) dargestellt. Zunächst werden die grundlegende Idee und die wichtigsten Resultate der Arbeit präsentiert (2.4.1). Dann erfolgt eine kurze Einordnung in die angrenzende Literatur (2.4.2). Daraufhin wird die Differenzenbildung als Methode zur Identifikation marginaler Effekte in nichtseparablen Funktionen vorgestellt (2.4.3). Dabei werden die wesentlichen Annahmen aufgezeigt, bevor mit Theorem 1 das wichtigste Ergebnis der Arbeit präsentiert wird. Dieses ermöglicht die Identifikation der sogenannten lokalen durchschnittlichen strukturellen Ableitungen. Zahlreiche Erweiterungen sind möglich. Einige davon werden diskutiert, bevor dann der Rahmen auf das binäre Modell erweitert wird und auch hierzu die wichtigsten Annahmen und Resultate dargestellt werden.

Angemerkt werden muss, dass die Arbeit von Hoderlein und White (2009) eine hoch komplexe Forschungsarbeit darstellt, die an die Grenzen der aktuellen ökonometrischen Forschung vordringt. Aufgrund der Schwierigkeit der Materie und einer

notwendigen Begrenzung des Umfangs, werden in dieser Diplomarbeit lediglich die wichtigsten Aspekte behandelt. Hierzu gehören wesentliche Annahmen, Ergebnisse und Erweiterungen. Tiefergehende Erklärungen oder Herleitungen sind ausdrücklich nicht Kern der Arbeit. Hierzu wird auf die Originalpublikation oder auf entsprechende weiterführende Literatur verwiesen.

Der eigentliche Fokus der Arbeit liegt auf der Anwendung der empirisch verwertbaren Aspekte der Arbeit von Hoderlein und White (2009). Zur Einführung in den Empirieteil erfolgen in Abschnitt 3.1 zunächst einige wichtige Hintergrundinformationen zur US-Immobilienkrise, die sich als Hauptauslöser der späteren globalen Wirtschafts- und Finanzkrise herausgestellt hat (3.1.1). Als nächstes werden einige der bisherigen Ansätze zur Erfassung von Immobilienmarkt-Risiken abgehandelt (3.1.2). Hierzu gehört beispielsweise der Case-Shiller-Index. Anschließend wird auf die Rolle der Haushalte und deren Verschuldung eingegangen (3.1.3). Dann wird der Beitrag dieser Diplomarbeit erläutert, der im Wesentlichen in einer neuartigen ökonometrischen Anwendung für deutsche Haushaltsdaten besteht (3.1.4).

Daraufhin wird in Abschnitt 3.2 der empirische Teil der Arbeit von Hoderlein und White (2009) dargestellt. Eingegangen wird auf die ökonometrische Umsetzung (3.2.1) und die Resultate (3.2.2).

Anschließend wird in Abschnitt 3.3 ausführlich die Untersuchung mit den deutschen Daten dargestellt. Erläutert werden das Sozio-ökonomische Panel als Datengrundlage (3.3.1), die Aufbereitung der Daten (3.3.2), die konkrete Schätzung (3.3.3) und die Ergebnisse (3.3.4). Diese können als mikroökonometrisch fundierten Erklärungsansatz dafür betrachtet werden, weshalb in Deutschland bisher keine allgemeine Immobilienkrise zu verzeichnen war.

Im Fazit (Abschnitt 4) erfolgt dann eine kurze Zusammenfassung der wichtigsten Aspekte und ein knapper Ausblick.

2. Theorieteil

2.1. Binäre Regression

Die folgenden Ausführungen basieren auf Wooldridge (2002), S. 574-587, Wooldridge (2006), S. 453-481 und Fahrmeir (2009), S. 189-197.

Liegt eine binäre abhängige Variable vor, so kann diese lediglich die Werte null und eins annehmen. Dargestellt werden kann dieser Fall mit dem linearen Wahrscheinlichkeitsmodell:

$$P(y = 1|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k. \quad (1)$$

Die Wahrscheinlichkeit dafür, dass die abhängige Variable gleich eins ist, wird als lineare Funktion der Variablen \mathbf{x} (bzw. x_1, \dots, x_k) modelliert. Bezeichnet wird dies auch als lineares Antwortmodell. $P(y = 1|\mathbf{x})$ wird als Antwortwahrscheinlichkeit bezeichnet. Da die Antwortwahrscheinlichkeit linear in den Parametern ist, misst der Koeffizient β_j die Änderung der Antwortwahrscheinlichkeit wenn sich x_j um eine Einheit ändert und dabei alle anderen Variablen konstant gehalten werden.

Zwar ist dieses Modell einfach zu handhaben, doch hat es zumindest zwei gewichtige Nachteile. Zum Einen können die resultierenden Antwortwahrscheinlichkeiten auch außerhalb des Intervalls $[0,1]$ liegen, wenn beliebige Werte für die Variablen \mathbf{x} eingesetzt werden. Zum Anderen sind die partiellen marginalen Effekte der Variablen aufgrund der Linearität in den Parametern immer konstant, egal ob kleine oder große Werte für die unabhängigen Variablen eingesetzt werden.

Die Überwindung dieser Schwächen ermöglichen Logit- und Probitmodelle. Angenommen sei eine Klasse von binären Antwort-Modellen:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = G(\beta_0 + \mathbf{x}\beta), \quad (2)$$

wobei G eine Funktion darstellt, die Werte strikt zwischen null und eins annimmt: $0 < G(z) < 1$ für alle reellen Zahlen z . Damit kann die Antwortvariable auch nur Werte zwischen null und eins annehmen. $\mathbf{x}\beta$ steht für $\beta_1 x_1 + \dots + \beta_k x_k$.

Im Logitmodell wird für G die logistische Funktion verwendet:

$$G(z) = \exp(z)/[1 + \exp(z)] = \Lambda(z). \quad (3)$$

Diese liegt wie gewünscht zwischen null und eins und entspricht der kumulativen Verteilungsfunktion für eine logistische Zufallsvariable.

Im Probitmodell wird für G die Verteilungsfunktion der Standardnormalverteilung verwendet. Diese kann ausgedrückt werden als:

$$G(z) = \Phi(z) \equiv \int_{-\infty}^z \phi(v) dv, \quad (4)$$

wobei $\phi(z)$ die Dichtefunktion der Standardnormalverteilung ist:

$$\phi(z) = (2\Pi)^{-1/2} \exp(-z^2/2). \quad (5)$$

Auch hier sichert $G(\cdot)$ ab, dass $P(y = 1|\mathbf{x})$ strikt zwischen null und eins liegt, egal welche Werte für x_j eingesetzt werden.

Logit- und Probitmodell sind einander sehr ähnlich, da G in beiden Fällen streng monoton wachsend ist und eine fast identische Form hat. Von Nachteil ist bei beiden Verfahren leider, dass die ermittelten Koeffizienten, im Gegensatz zu denen im linearen Wahrscheinlichkeitsmodell, nicht direkt interpretierbar sind. Die Funktionen für G sind nichtlinearer Natur. Zwar ist das Vorzeichen der standardmäßig ausgewiesenen Koeffizienten und der interessierenden partiellen Effekte dasselbe. Außerdem ist der relative Effekt zweier stetiger Zufallsvariablen unabhängig von den Werten der \mathbf{x} bestimmbar. Das Ausmaß der partiellen Effekte hängt jedoch von den für \mathbf{x} eingesetzten Werten ab.

Prinzipiell können die partiellen Effekte über die jeweiligen Ableitungen gewonnen werden:

$$\frac{\delta P(y = 1|\mathbf{x})}{\delta x_j} = g(\beta_0 + \mathbf{x}\beta)\beta_j, \quad (6)$$

wobei $g(z) \equiv \frac{dG}{dz}(z)$. Da G die kumulative Verteilungsfunktion einer stetigen Zufallsvariable darstellt, entspricht g einer Wahrscheinlichkeitsdichte.

Stellt x_1 eine binäre erklärende Variable dar, dann ist der partielle Effekt der Änderung von x_1 von null zu eins, unter der Voraussetzung dass alle anderen Einflussfaktoren konstant bleiben, einfach:

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \dots + \beta_k x_k). \quad (7)$$

Wiederum hängt der Effekt von den für die anderen \mathbf{x}_j eingesetzten Werten ab.

Geschätzt werden Logit- und Probitmodelle über das Maximum-Likelihood (ML)-Verfahren. Um den ML-Schätzer, bedingt für die erklärenden Variablen, zu erhalten, benötigt man die Dichte von y_i für gegebene \mathbf{x}_i . Dies kann ausgedrückt werden als

$$f(y|\mathbf{x}_i; \beta) = [G(\mathbf{x}_i\beta)]^y [1 - G(\mathbf{x}_i\beta)]^{1-y}, y = 0, 1, \quad (8)$$

wobei der Achsenabschnitt der Einfachheit halber in den Vektor \mathbf{x}_i aufgenommen wird. Für $y = 1$ erhält man also $G(\mathbf{x}_i\beta)$ und für $y = 0$ erhält man $1 - G(\mathbf{x}_i\beta)$.

Die Log-Likelihood-Funktion für die Beobachtung i ist eine Funktion der Parameter und der Daten \mathbf{x}_i, y_i und wird erhalten indem man Gleichung 8 logarithmiert:

$$l_i(\beta) = y_i \log[G(\mathbf{x}_i\beta)] + (1 - y_i) \log[1 - G(\mathbf{x}_i\beta)]. \quad (9)$$

Da G sowohl im Logit-, als auch im Probitmodell strikt zwischen null und eins liegt, ist $l_i(\beta)$ für alle Werte von β definiert.

Die Log-Likelihood für eine Stichprobe der Größe n erhält man dann, indem man Gleichung 9 über sämtliche Beobachtungen hinweg aufsummiert: $L(\beta) = \sum_{i=1}^n l_i(\beta)$. Der ML-Schätzer von β , bezeichnet durch $\hat{\beta}$, maximiert diese Log-Likelihood. Entspricht $G(\cdot)$ der logistischen kumulativen Verteilungsfunktion, dann ist $\hat{\beta}$ der Logit-Schätzer, entspricht $G(\cdot)$ der kumulativen Verteilungsfunktion der Standardnormalverteilung, dann ist $\hat{\beta}$ der Probit-Schätzer. Diese Schätzer sind unter sehr allgemeinen Bedingungen konsistent, asymptotisch normalverteilt und effizient. Sobald die Standardfehler berechnet sind, können die üblichen Tests durchgeführt werden. Für die in Binärmodellen oft problematische Heteroskedastizität wird in der Regel automatisch kontrolliert.

Da die meisten Schritte von der Mehrzahl der Statistik-Softwares automatisch durchgeführt werden, liegt das Hauptproblem in der Regel in der Berechnung der partiellen Effekte. Da diese wie bereits erwähnt von den eingesetzten Werten für die \mathbf{x} abhängt, gibt es prinzipiell beliebige Möglichkeiten zu deren Bestimmung. In der Regel werden jedoch entweder durchschnittliche marginale Effekte oder marginale Effekte an den Mittelwerten ausgewiesen. Letztere kann man sich als marginale Effekte vorstellen, die für eine in allen Variablen durchschnittliche Beobachtungseinheit gelten. Erstere fassen den Durchschnitt aller individuellen marginalen Effekte zusammen. Für beide Varianten gibt es in den meisten Softwares Standardbefehle.

Zur Anwendung kommen Probit- und Logitschätzung auch im Kontext von Paneldaten und nonparametrischen Regressionen. Näheres dazu in den folgenden Abschnitten.

2.2. Paneldaten und Fixed Effects

Grundlegend für diesen Abschnitt sind Wooldridge (2002), S. 247-291, S. 483-490 und Wooldridge (2006), S. 481-536.

Wie bereits einleitend erwähnt, werden in Paneldatenmodellen Beobachtungen für bestimmte Einheiten (zum Beispiel Individuen) über mehrere Perioden erfasst. Formal kann dies folgendermaßen ausgedrückt werden:

$$y_t = \mathbf{x}_t\beta + u_t \quad t = 1, 2, \dots, T. \quad (10)$$

Es handelt sich hierbei um ein einfaches statisches, lineares Paneldatenmodell. \mathbf{x}_t stellt einen $1 \times K$ Vektor für die Perioden $t = 1, 2, \dots, T$ dar, β ist ein $K \times 1$ Vektor und y_t ist ein Skalar.

Da bei Hoderlein und White (2009) ein zweiperiodiges Modell im Fokus steht, wird hier zum Einstieg ein lineares Modell, ebenfalls für den Spezialfall mit $T = 2$ erläutert. Liegt nur eine erklärende Variable vor, so kann das Modell folgendermaßen ausgedrückt werden:

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + c_i + u_{it} \quad t = 1, 2. \quad (11)$$

Dabei steht i für die jeweilige Beobachtungseinheit (Individuum 1,2,...), während t für die entsprechende Periode steht. $d2_t$ stellt eine Dummy-Variable dar, die den Wert null annimmt für $t = 1$, eins für $t = 2$ und sich über die Beobachtungseinheiten hinweg nicht ändert. Für $t = 1$ lautet der Achsenabschnitt daher β_0 , für $t = 2$ dagegen $\beta_0 + \delta_0$. u_{it} stellt den idiosynkratischen bzw. über die Zeit variablen Fehler dar, während die Variable c_i alle unbeobachtbaren, zeitkonstanten Faktoren enthält, die einen Einfluss auf y_{it} haben. c_i erfasst damit die unbeobachtbare Heterogenität zwischen den Beobachtungseinheiten. Diese kann beispielsweise das Resultat unterschiedlicher Präferenzen von Individuen sein. Allgemein wird c_i als unbeobachteter Effekt bezeichnet. Alternativ wird der Ausdruck Fixed Effect verwendet. Dieser bringt zum Ausdruck, dass c_i sich im Zeitverlauf nicht ändert. Das gesamte durch Gleichung 11 aufgestellte Modell wird Fixed Effects (FE) Modell oder Modell für unbeobachtbare Effekte genannt.

Für die Berechnung des interessierenden Koeffizienten β_1 kommen mehrere Verfahren in Betracht. Eine gewöhnliche Kleinstquadrateschätzung mit den gepoolten Daten liefert jedoch nicht die gewünschten Ergebnisse. Das Hauptproblem stellt die soge-

nannte Heterogenitätsverzerrung dar. Diese tritt auf sobald der recht wahrscheinliche Fall auftritt, dass c_i mit x_i korreliert ist. Die Voraussetzung, dass lediglich u_{it} und x_{it} unkorreliert sind, ist nicht ausreichend.

Gelöst werden kann dieses Problem anhand der Methode des First Differencing (FD) bzw. der Differenzenbildung. Die Gleichungen für die beiden Perioden können folgendermaßen ausgedrückt werden:

$$\begin{aligned} y_{i2} &= \beta_0 + \delta_0 + \beta_1 x_{i2} + c_i + u_{i2} \quad t = 2 \\ y_{i1} &= \beta_0 + \beta_1 x_{i1} + c_i + u_{i1} \quad t = 1. \end{aligned} \tag{12}$$

$\beta_0 + \delta_0$ stellt dabei den Achsenabschnitt für die zweite Periode dar.

Zieht man nun die erste Gleichung von der zweiten Gleichung ab, so erhält man

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i. \tag{13}$$

Der unbeobachtbare Effekt c_i kommt in dieser Gleichung nicht mehr vor. Partielle Effekte sind nun leider nicht mehr ausgehend von den absoluten Werten bestimmbar. β_1 zeigt nun wie sich Änderungen der erklärenden Variable auf Änderungen in der abhängigen Variable auswirken (von $t = 1$ auf $t = 2$). Sind Δu_i und Δx_i miteinander unkorreliert, so ist der Schätzer für β_1 konsistent. Dies ist dann erfüllt, wenn der idiosynkratische Fehler u_{it} in beiden Perioden mit den erklärenden Variablen unkorreliert ist. Von praktischer Bedeutung ist zudem, dass Δx_i über ausreichend Variation verfügt. Ist dies nicht gegeben, so resultieren große Standardfehler für $\hat{\beta}_1$. Gilt $x_i = 0$ für alle i , so ist die Schätzung der Gleichung nicht möglich. Bei Heteroskedastizität sind die gängigen Korrekturmethoden anwendbar.

Alternativ zur Differenzenbildung bietet sich die FE-Transformation an. Betrachtet sei hier wieder ein Modell mit lediglich einer erklärenden Variable:

$$y_{it} = \beta_1 x_{it} + c_i + u_{it}, \quad t = 1, 2, \dots, T. \tag{14}$$

Nun wird für alle i der Durchschnitt über die Zeit gebildet. Man erhält:

$$\bar{y}_i = \beta_1 \bar{x}_i + c_i + \bar{u}_i. \tag{15}$$

Da c_i im Zeitverlauf fix ist, entspricht der Mittelwert c_i . Zieht man nun Gleichung 15 von Gleichung 14 ab, so erhält man die um die über die Mittelwerte nach der Zeit bereinigten Daten:

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i, \quad t = 1, 2, \dots, T, \quad (16)$$

bzw.

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it} \quad t = 1, 2, \dots, T. \quad (17)$$

Auch hier verschwindet die unbeobachtbare Komponente c_i . Gleichung 17 kann über ein gepooltes Kleinstquadratverfahren geschätzt werden. Der auf diesen um die Mittelwerte nach der Zeit bereinigten Daten basierende Schätzer wird FE-Schätzer genannt. Zu beachten ist, dass für die Unverzerrtheit des Schätzers die strikte Exogenität der erklärenden Variablen notwendig ist. c_i darf jedoch nach wie vor mit den erklärenden Variablen beliebig korrelieren. Sämtliche im Zeitverlauf konstante Variablen fallen im Rahmen der FE-Schätzung heraus.

Im Fall von $T = 2$ sind die FE- und FD- Schätzer sowie deren Teststatistiken identisch. Ist $T \geq 2$ so unterscheiden sich FE- und FD-Schätzer. Angemerkt sei, dass bei seriell unkorrelierten u_{it} die FE-Schätzung effizienter ist als die FD-Schätzung. Da das Modell für unbeobachtbare Variablen normalerweise für den Fall seriell unkorrelierter Fehler aufgestellt wird, wird der FE-Schätzer in der Praxis etwas häufiger verwendet. Für den Fall, dass unbeobachtete Variablen, die nicht mit den erklärenden Variablen korreliert sind vorliegen, kann auf das sogenannte Random Effects Modell zurückgegriffen werden.

Erweitert werden kann das Modell für unbeobachtete Effekte auch auf den Logit- und den Probitrahmen. Für den Probitrahmen lautet die wesentliche Annahme:

$$P(y_{it} = 1 | \mathbf{x}_i, c_i) = P(y_{it} = 1 | \mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\beta + c_i), \quad t = 1, \dots, T. \quad (18)$$

c_i sei wieder der unbeobachtete Effekt und \mathbf{x}_i enthalte \mathbf{x}_{it} für alle t . Der linke Teil der Gleichung besagt, dass \mathbf{x}_{it} strikt exogen ist, sobald c_i berücksichtigt wird. Es erscheint dann nur noch \mathbf{x}_{it} in der Antwortwahrscheinlichkeit zum Zeitpunkt t . Dies schließt zeitlich verzögerte abhängige Variablen als Regressoren aus. Der zweite Teil der Gleichung zeigt die Standard Probit-Annahme, dass c_i additiv im Index innerhalb von $\Phi(\cdot)$ vorkommt. Eine zusätzliche Annahme ist, dass die y_{i1}, \dots, y_{iT} unab-

hängig sind bedingt für (\mathbf{x}_i, c_i) . Schließlich kann unter diesen Standardannahmen die Dichte von y_{i1}, \dots, y_{iT} bedingt für (\mathbf{x}_i, c_i) hergeleitet werden:

$$f(y_1, \dots, y_T | \mathbf{x}_i, c_i; \beta) = \prod_{t=1}^T f(y_t | \mathbf{x}_{it}, c_i; \beta), \quad (19)$$

wobei $f(y_t | \mathbf{x}_t, c; \beta) = \Phi(\mathbf{x}_t \beta + c)^{y_t} [1 - \Phi(\mathbf{x}_t \beta + c)]^{1-y_t}$. Für die Schätzung der interessierenden Effekte wäre es wünschenswert, möglichst keine Beschränkung bezüglich der Beziehung zwischen c_i und \mathbf{x}_{it} anzunehmen. In einer Fixed Effects Probit-Analyse würden die c_i als Parameter zusammen mit den β geschätzt werden. Die Log-Likelihood-Funktion ist $\sum_{i=1}^N l_i(c_i, \beta)$, wobei $l_i(c_i, \beta)$ den Logarithmus von Gleichung 19 bewertet zu y_{it} darstellt. Das Rechenverfahren ist jedoch leider nicht nur komplex sondern führt zum „Incidental Parameters Problem“.³ Die Schätzung liefert inkonsistente Schätzer für β . Im linearen Fall war dies nicht der Fall. Die Frage ob eine Verteilung von c_i gegeben \mathbf{x}_i spezifiziert werden muss, ist daher für viele mikroökonomische Anwendungen ein wichtiger Punkt.

Im Probit-Rahmen seien daher die c_i als unbeobachtbare Zufallsvariablen behandelt, ähnlich wie in linearen Modellen. Sie werden quasi gemeinsam mit $\mathbf{x}_i, \mathbf{y}_i$ „gezogen“ (Die Bezeichnung FE Probit ist in dieser Hinsicht irritierend). Um eine konsistente Schätzung der partiellen Effekte zu ermöglichen, müssen für das Probit-Modell unbeobachteter Variablen Annahmen für die Beziehung zwischen c_i zu \mathbf{x}_i getroffen werden. Das traditionelle Random Effects Probit Modell fügt zu den obigen Standardannahmen noch folgende Annahme hinzu:

$$c_i | \mathbf{x}_i \approx \text{Normal}(0, \sigma_c^2). \quad (20)$$

Dies impliziert, dass c_i und \mathbf{x}_i unabhängig sind und dass c_i normalverteilt ist. Lediglich anzunehmen, dass c_i und \mathbf{x}_i unkorreliert sind oder dass $E(c_i | \mathbf{x}_i) = 0$ reicht also nicht aus. Ist im Modell ein Achsenabschnitt enthalten, so kann $E(c_i) = 0$ angenommen werden, ohne dass ein Verlust entsteht.

Eine konsistente Schätzung von β bedeutet, dass die partiellen Effekte der Elemente von \mathbf{x}_t auf die Antwortwahrscheinlichkeit $P(y_t = 1 | \mathbf{x}_t, c)$ korrekt bestimmt sind. c stellt den Durchschnittswert der c_i dar (für $E(c_i) = 0$ ist $c = 0$). Die relativen Effekte zweier Elemente von \mathbf{x}_t können auch für beliebige Werte von c bestimmt werden, da diese nicht von c abhängen. Aufgrund der Nützlichkeit, werden jedoch in der Regel durchschnittliche partielle Effekte ausgewiesen. Da gilt $c_i \approx \text{Normal}$

³ Näheres zu diesem Problem findet sich beispielsweise bei Lancaster (2000).

$(0, \sigma_c^2)$, sind die durchschnittlichen partiellen Effekte für eine stetige Variable x_{tj} : $[\beta_j/(1 + \sigma_c^2); 1/2]\phi[\mathbf{x}_t\beta/(1 + \sigma_c^2)^{1/2}]$. Deshalb muss lediglich $\beta_c \equiv \beta/(1 + \sigma_c^2)^{1/2}$ geschätzt werden, um die durchschnittlichen partiellen Effekte für stetige oder diskrete erklärende Variablen zu berechnen.

Sind die obigen Annahmen erfüllt, so ist ein bedingter ML-Ansatz zur Schätzung von β und σ_c^2 verfügbar. Da die c_i nicht beobachtet werden können, können diese auch nicht in der Likelihood-Funktion auftauchen. Um die gemeinsame Verteilung von (y_{i1}, \dots, y_{iT}) bedingt für \mathbf{x}_i zu finden, muss der Einfluss der c_i herausintegriert werden. Da c_i eine Normalverteilung $(0, \sigma_c^2)$ hat, gilt:

$$f(y_1, \dots, y_T | \mathbf{x}_i; \phi) = \int_{-\infty}^{\infty} \left[\prod_{t=1}^T f(y_t | \mathbf{x}_{it}, c; \beta) \right] (1/\sigma_c) \phi(c/\phi_c) dc, \quad (21)$$

wobei $f(y_t | \mathbf{x}_t, c; \beta) = \Phi(\mathbf{x}_t\beta + c)^{y_t} [1 - \Phi(\mathbf{x}_t\beta + c)]^{1-y_t}$ und θ enthält β und σ_c^2 . Einsetzen von y_{it} für alle t und Logarithmieren von Gleichung 21 liefert die bedingte Log-Likelihood $l_i(\theta)$ für jedes i . Die Log-Likelihood-Funktion für die gesamte Stichprobe der Größe N kann maximiert werden nach β und σ_c^2 um \sqrt{N} -konsistente asymptotisch normale Schätzer zu erhalten. Ein Verfahren zur Approximation des obigen Integrals wird von Butler und Moffit (1982) beschrieben. Der in diesem Zusammenhang berechnete bedingte ML-Schätzer wird als Random Effects Schätzer bezeichnet. Da β und σ_c^2 geschätzt werden können, sind die partiellen Effekte bei $c = 0$ sowie die durchschnittlichen partiellen Effekte ebenfalls schätzbar. Da die Varianz des zusammengesetzten Fehlers im Modell mit latenter Variable eins ist, wird die relative Wichtigkeit des unbeobachteten Effekts gemessen als $\rho = \sigma_c^2/(\sigma_c^2 + 1)$, was der Korrelation zwischen dem zusammengesetzten latenten Fehler $(c_i + e_{it})$ über zwei beliebige Perioden hinweg entspricht. Dieses $\hat{\rho}$ und sein Standardfehler werden von vielen Softwares standardmäßig ausgewiesen. Damit kann relativ unkompliziert auf das Vorhandensein eines unbeobachteten Effekts getestet werden.

2.3. Semi- und nonparametrische Verfahren

Grundlegend für diesen Abschnitt sind Cameron (2005), S. 294-334, Myoung-jae (2010), S. 303-362, Myoung-jae (2006), S. 123-166 sowie Fahrmeir (2009), S. 291-398.

2.3.1. Überblick

Beim Großteil der in der Praxis verwendeten Regressionsverfahren handelt es sich um parametrische Verfahren. Diese Verfahren sind einfach zu handhaben und leicht zu interpretieren: Partielle Effekte der erklärenden Variablen werden durch die jeweiligen Regressionskoeffizienten ausgedrückt. Unterstellt wird Linearität in den Parametern.

Tatsächlich können die Beziehungen zwischen den Variablen auch von beliebiger nichtlinearer Natur sein. Dann sind nonparametrische Verfahren von Vorteil. Gemeint sind Verfahren, die es ermöglichen, keine Annahmen über die Wahrscheinlichkeitsverteilung der abhängigen Variable zu treffen (siehe Manski (1975)) und die mathematische Form des Zusammenhangs zwischen den unabhängigen Variablen und der abhängigen Variable offen zu lassen. Der Einfluss der unabhängigen Variablen wird durch (innerhalb gewisser Grenzen) beliebige Funktionen dieser Variablen dargestellt. Es resultieren daher keine einzelnen Regressionskoeffizienten, sondern geschätzte Funktionswerte für die in den Daten vorkommenden Werte der unabhängigen Variablen.

Wird nun doch eine Regressionsfunktion spezifiziert, die Verteilung der abhängigen Variable jedoch offen gelassen, so wird das Verfahren als semiparametrisch im eigentlichen Sinn bezeichnet. Zur Bestimmung der Schätzer werden dann keine nonparametrischen Techniken benötigt. Sobald diese jedoch zumindest für einen Teil der Variablen eingesetzt werden, wird das Verfahren als semi-nonparametrisch bezeichnet.

Der Hauptgrund für den Einsatz dieser Verfahren besteht darin, dass multivariate Regressionen oft vom „Fluch der Dimensionalität“ betroffen sind: Die Konvergenzrate der geschätzten Funktion verringert sich mit der Zahl der Regressoren, da durch die Einteilung der Daten in verschiedene Dimensionen bzw. Schichten zu wenige Datenpunkte in den einzelnen Schichten verbleiben. Semiparametrische Verfahren schaffen hier Abhilfe, da durch die parametrische Spezifikation bestimmter Regressoren die Zahl der Dimensionen reduziert werden kann.

Zu den Begrifflichkeiten muss angemerkt werden, dass der Sprachgebrauch leider nicht ganz einheitlich ist. Viele Autoren verstehen unter semiparametrischen Verfahren nicht nur die semiparametrischen Verfahren im eigentlichen Sinn, sondern auch die semi-nonparametrischen Verfahren. Da dies auch bei Hoderlein und White (2009) der Fall ist, sind im Folgenden unter semiparametrischen Verfahren in der Regel auch die semi-nonparametrischen Verfahren zu verstehen. Geht es explizit um den Einsatz semi-nonparametrischer Verfahren, so werden diese auch so bezeichnet.

Der wichtigste Vorteil der semi- und nonparametrischen Verfahren ist deren Flexibilität. Die in den Daten liegenden Zusammenhänge können theoretisch exakt von entsprechenden Modellen wiedergegeben werden. Es kann zwar zu einem Overfitting kommen. Dieses kann jedoch mit Hilfe der Kreuzvalidierung gelöst werden: Der ursprüngliche Datensatz wird in einen Trainings- und einen Validierungsdatsatz unterteilt. Dann wird überprüft ob der für den Trainingsdatensatz ermittelte Zusammenhang auch gleichermaßen für den Validierungsdatsatz gilt. Ergibt sich ein Widerspruch, so wird die Regressionsfunktion so verändert, dass der Zusammenhang eine größere Allgemeingültigkeit erlangt.

Die Anwendung semi- und nonparametrischer Verfahren findet sich vor allem im Rahmen der explorativen Datenanalyse: Neue Theorien werden getestet. Bestehende Theorien werden verfeinert. Oft liefern nonparametrische Verfahren auch die nötigen Anhaltspunkte für die Formulierung von letztlich doch parametrischen Hypothesen. Umgekehrt können sie auch zur Überprüfung von Vorhersagen parametrischer Verfahren verwendet werden. Des Weiteren sind die Verfahren im Zusammenhang mit Kontrollvariablen von großem Nutzen. Wird deren Einfluss nichtparametrisch bestimmt, so können unter Umständen die Effekte der interessierenden Variablen genauer bestimmt werden.

2.3.2. Kernel-Regression

Die konkrete Schätzung der Regressionsfunktionen erfolgt bei den nonparametrischen Verfahren über sogenannte Streudiagramm-Glätter (Scatterplot-Smoother). Verwendung finden beispielsweise das gleitende Mittel oder die lokale polynomiale Regression. Besonders wichtig sind die sogenannten Kernel-Methoden. Da diese auch in der Arbeit von Hoderlein und White (2009) Verwendung finden, wird hierzu auf die wichtigsten Aspekte eingegangen.

Angenommen man habe N Beobachtungen x_i mit $i = 1, \dots, N$ wobei x_i einen $k \times 1$ Vektor darstellt. Das Interesse liege auf $P(x \leq x_0) \equiv F(x_0)$. Dann kann $P(x \leq x_0)$ über die empirische Verteilungsfunktion $(1/N) \sum_{i=1}^N 1[x_i \leq x_0]$ geschätzt werden.⁴ Diese konvergiert zwar gegen $F(x_0)$, ist jedoch im Gegensatz zu $F(x_0)$ nicht differenzierbar. Das Differenzieren der empirischen Verteilungsfunktion stellt daher keine Option zur Bestimmung der Dichtefunktion $f(x_0)$ dar. Da jedoch gilt $f(x_0) = dF(x_0)/dx$, ist es möglich, $f(x_0)$ zu schätzen, indem man $dF(x_0)$ und dx approximiert:

h sei eine kleine, positive Zahl. Es gelte $dx \cong h^k$, dann ist

$$\begin{aligned} dF(x_0) &= P(x_0 < x < x_0 + h) \cong (1/N) \sum_{i=1}^N 1[x_0 < x_i < x_0 + h]. \\ &= (1/N) \sum_i (1/2) \cdot 1[x_0 - h < x_i < x_0 + h] \end{aligned} \quad (22)$$

Daher ist ein Dichteschätzer, der $dF(x_0)/dx$ approximiert:

$$\begin{aligned} &(1/(Nh^k)) \sum_i (1/2) \cdot 1[x_0 - h < x_i < x_0 + h] \\ &= (1/(Nh^k)) \sum_i (1/2) \cdot 1[-h < x_i - x_0 < h]. \end{aligned} \quad (23)$$

Damit diese Annäherung funktioniert, muss h möglichst klein sein. Ist h jedoch zu klein, dann gibt es keine Beobachtung mehr, die $-h < x_i - x_0 < h$ erfüllt. Deshalb kann $h \rightarrow +0$ nur dann angenommen werden, wenn gleichzeitig $N \rightarrow \infty$.

Sieht man die Indikatorfunktion in Gleichung 23 als eine Gewichtungsfunktion an, die nur das Gewicht 1 liefert, wenn x_i innerhalb der h -Distanz von x_0 liegt und ansonsten das Gewicht 0 liefert, dann kann die Schätzung in Gleichung 23 verallgemeinert werden:

$$\hat{f}(x_0) \equiv (1/(Nh^k)) \sum_i K((x_i - x_0)/h). \quad (24)$$

K stellt eine glatte Gewichtungsfunktion dar und wird als Kern bezeichnet. Insgesamt handelt es sich bei (24) um den von Rosenblatt (1956) eingeführten Kerndichteschätzer. Wählt man in (24) $K(z) = 1[-1 < z < 1]/2$ so resultiert wieder (23). Prinzipiell bleibt die Wahl des Kerns jedoch dem Anwender überlassen. Die Grundidee besteht darin, dass das Zentrum des Kerns auf jeden Datenpunkt gelegt wird und gezählt wird, wie viele weitere Datenpunkte sich in der nahen Umgebung

⁴ $1[\cdot]$ ist eine Indikatorfunktion:

$1(A) = 1$ falls A gilt

$1(A) = 0$ sonst.

befinden. Der Beitrag jedes einzelnen Punktes wird dann zu einem Gesamtschätzer zusammengefügt, wobei in der Regel nahe Punkte ein größeres Gewicht bekommen als entferntere Punkte. In Anlehnung an Lee (1996), verwendet man normalerweise Funktionen mit den folgenden Eigenschaften:

$$\begin{aligned}
& (i) \ K(z) \text{ ist symmetrisch um } 0 \text{ und stetig.} \\
& (ii) \ \int K(z)dz = 1, \int K(z)zdz = 0_k, \text{ und } \int |K(z)|dz < \infty. \\
& (iii) \ (a) \ K(z) = 0 \text{ für } |z| \geq z_0 \text{ für einige } z_0, \text{ oder} \\
& \quad (b) \ |z| \cdot K(z) \rightarrow 0 \text{ für } |z| \rightarrow \infty. \\
& (iv) \ K(z) = \prod_{j=1}^k L(z_j), \text{ wobei } L(i) \text{ bis (iii) erfüllt, und } \int t^2 L(t)dt = \kappa.
\end{aligned} \tag{25}$$

Fast alle in der Praxis verwendeten Kern-Funktionen erfüllen (i) bis (iii). Beispiele für Kern-Funktionen, die die Bedingungen in (25) für $k = 1$ erfüllen, sind:

$$\begin{aligned}
& (i) \ 1[|z| < 1]/2 : \text{uniformer Kern;} \\
& (ii) \ N(0, 1) \text{ Dichte } \phi(z); \\
& (iii) \ (3/4)(1 - z^2)1[|z| < 1] : \text{quadratischer Kern;} \\
& (iv) \ (15/16)(1 - z^2)^2 1[|z| < 1] : \text{quartischer Kern;}
\end{aligned} \tag{26}$$

Der uniforme Kern ist nicht glatt und wird daher selten benutzt (er liefert ein Histogramm). Die anderen drei Kerne werden dagegen häufig verwendet. Für $k > 1$ können auch Produkte eindimensionaler Kerne, wie $\prod_{j=1}^k \phi(z_j)$ verwendet werden. Ebenso sind multivariate Kerne wie die multivariate Normaldichte mit Varianz Ω denkbar:

$$K(z) = (2\pi)^{-k/2} \cdot |\Omega|^{-1/2} \cdot \exp\{-(1/2) \cdot z' \Omega^{-1} z\}. \tag{27}$$

Wichtig für die geschätzte Dichtefunktion ist, dass keine Verzerrung vorliegt und die Varianz möglichst gering gehalten wird. Punktweise Konsistenz liegt vor für $h \rightarrow 0$ und $Nh \rightarrow \infty$. Dies ist von Bedeutung für die Berechnung punktwiser Konfidenzbänder. Die Varianz geht gegen null, wenn $Nh^k \rightarrow \infty$. Die Verzerrung geht gegen null wenn $h \rightarrow 0$. Für asymptotische Unverzerrtheit in Verteilung muss gelten: $Nh^{k+4} \rightarrow 0$. Es lässt sich festhalten, dass wenn f zweifach stetig differenzierbar ist mit beschränkten Ableitungen und K (26)(i) und (ii) erfüllt, dass dann gilt

$\hat{f}(x_0) \stackrel{p}{=} f(x_0)$, wenn $Nh^k \rightarrow \infty$ und $h \rightarrow 0$, für $N \rightarrow \infty$. Eine noch bessere Eigenschaft stellt jedoch die uniforme Konsistenz dar:

$$\sup_{x_0} |\hat{f}(x_0) - f(x_0)| = o_p(1). \quad (28)$$

Ist diese gegeben, so konvergiert die maximale Abweichung von $\hat{f}(x_0)$ von $f(x_0)$ gegen null. Es ist dann abgesichert, dass sich nicht nur der „Graph der Grenzwerte“, sondern auch der „grenzwertige Graph“ möglichst eng an den gegebenen Zusammenhang annähert. Unter bestimmten Voraussetzungen kann gezeigt werden, dass wenn $Nh^k/\ln(N) \rightarrow \infty$ und $h \rightarrow 0$ für $N \rightarrow \infty$, dass $\sup_x |\hat{f}(x) - f(x)| \stackrel{as}{=} 0$ und folglich $\hat{f}(x) \stackrel{as}{=} f(x)$ gilt. Die Rate $Nh^k/\ln(N) \rightarrow \infty$ verlangt damit ein etwas größeres h , als dasjenige bei $Nh^k \rightarrow \infty$ (im Fall punktwiser Konvergenz). Ein zu kleines h macht $\hat{f}(x_0)$ mehr gezackt. Dies wirkt der angestrebten Uniformität entgegen.

Insgesamt spielt die Wahl des Glättungsparameters bzw. der Bandbreite h eine große Rolle. Ein kleiner Glättungsparameter sorgt für eine geringe Verzerrung. Werden nur wenige Beobachtungen zur Berechnung von $\hat{f}(x_0)$ verwendet, so kann $\hat{f}(x_0)$ zu sehr gezackt sein, die Varianz von $\hat{f}(x_0)$ wird zu groß. Ein großer Glättungsparameter kann dagegen dazu führen, dass $\hat{f}(x_0)$ nur geringe Änderungen aufweist, wenn x_0 sich ändert. Im Extremfall stellt $\hat{f}(x_0)$ dann eine Konstante dar, die gar nicht auf Änderungen von x_0 reagiert. Ein gut gewählter Glättungsparameter muss daher zwischen den Extremen liegen.

Ein Verfahren zur Bestimmung eines optimalen Glättungsparameters h stellt die Kleinstquadrat-Kreuzvalidierung dar. In diesem Daten-basierten Verfahren wird ein Schätzer für den integrierten quadrierten Fehler um x_0 nach h minimiert. Das Verfahren wird von den meisten Software-Paketen automatisch ausgeführt. Es hat jedoch die Tendenz, insbesondere bei kleinen h , mehrere lokale Minima zu haben. Ein Undersmoothing ist daher recht wahrscheinlich. Zu beachten ist außerdem, dass das ermittelte, optimale h nur für den gegebenen Datensatz optimal ist. Für andere Datensätze kann wiederum ein anderes h optimal sein. Trial-and-Error-Verfahren liefern trotz allem oft noch brauchbare Informationen.

Die Wahl des optimalen Kerns ist ebenfalls von großer Bedeutung. Sie stellt sich als schwieriger als die Wahl von h dar, da K eine Funktion ist. Über dynamische Optimierung kommt man jedoch auf den Epanechnikov-Kern:

$$K(Z) = \{3/(4\sqrt{5})\} \cdot (1 - 0.2z^2) \cdot 1[|z| \leq \sqrt{5}]. \quad (29)$$

Dieser ist der effizienteste aller Kerne und wird daher in der Regel standardmäßig verwendet. Im Vergleich zur Wahl von h muss jedoch angemerkt werden, dass die Wahl des Kerns nur eine untergeordnete Rolle spielt. Obwohl der Epanechnikov-Kern den mittleren integrierten quadrierten Fehler minimiert, hat der normale uniforme Kern eine im Vergleich 93%ige Effizienz. Die Unterschiede unter den geläufigen Kernen sind somit eher gering.

Nachdem die Bestimmung der Dichtefunktion dargestellt wurde, wird nun auf die nonparametrische Bestimmung der Regressionsfunktion $E(y|x) \equiv r(x)$ (mit $y = r(x) + u$ und $E(u) = 0$) eingegangen. Da $E(y|x)$ dem Ausdruck $\int \{f(y, x)/f(x)\}y dy$ entspricht, kann $E(y|x)$ bestimmt werden sobald die Dichtefunktion ermittelt ist. Die Kernel-Schätzung stellt hierzu wieder die Methode dar.

Liegen für ein gegebenes x_0 mehrere y_i 's vor, so kann $E(y|x_0)$ über den Stichprobendurchschnitt dieser Beobachtungen ermittelt werden. Verfügt man jedoch über eine Zufallsvariable x , dann ist dies nicht möglich, denn bei jedem erneuten Ziehen bekommt man andere x_i 's. Nimmt man jedoch an dass $E(y|x) = r(x)$ stetig in x ist und die x_i 's der entsprechenden y_i 's nahe an x_0 liegen, dann kann dies behandelt werden als (y_i, x'_0) . Dies bedeutet, dass nun für ein gegebenes x_0 multiple Beobachtungen vorliegen können und $E(y|x_0)$ eben doch über den Stichprobendurchschnitt ermittelt werden kann. Allgemein erfolgt die Schätzung über ein gewichtetes Mittel $\sum_i w_i y_i$, wobei w_i groß ist, wenn x_i nahe an x_0 und klein ist, wenn x_i weiter entfernt von x_0 liegt. Außerdem gilt $\sum_i w_i = 1$. Für verschiedene Gewichtungen kommen wieder unterschiedliche Kerne in Frage. Ein möglicher Schätzer für $r(x_0)$ ist

$$\sum_i K((x_i - x_0)/h) y_i. \quad (30)$$

Um sicherzustellen, dass sich die Gewichte auf eins addieren, normalisiert man obige Gleichung und erhält

$$\begin{aligned} r_N(x_0) &= \sum_i K((x_i - x_0)/h) y_i / \sum_i K((x_i - x_0)/h) \\ &= \sum_i \{K((x_i - x_0)/h) / \sum_i K((x_i - x_0)/h)\} y_i = \sum_i w_i y_i \\ &= (Nh^k)^{-1} \sum_i K((x_i - x_0)/h) y_i / (Nh^k)^{-1} \sum_i K((x_i - x_0)/h) \\ &= g_N(x_0) / f_N(x_0), \end{aligned} \quad (31)$$

wobei

$$g_N(x_0) \equiv (Nh^k)^{-1} \sum_i K((x_i - x_0)/h) y_i \quad (32)$$

und

$$f_N(x_0) \equiv (Nh^k)^{-1} \sum_i K((x_i - x_0)/h). \quad (33)$$

Es handelt sich um den sogenannten Nadaraya-Watson Kernel-Schätzer. Es gilt $f_N(x_0) \stackrel{p}{=} f(x_0)$. Damit $r_N(x_0) \stackrel{p}{=} r(x_0)$ ebenfalls gilt, benötigt man:

$$g_N(x_0) \stackrel{p}{=} g(x_0) \equiv r(x_0) \cdot f(x_0). \quad (34)$$

Verständlich wird dies, wenn x diskret ist. $E(y|x_0)$ ist dann schätzbar über den Stichprobendurchschnitt der y_i 's mit $x_i = x_0$:

$$\begin{aligned} & \sum_i 1[x_i = x_0] y_i / \sum_i 1[x_i = x_0] \\ &= (1/N) \sum_i 1[x_i = x_0] y_i / (1/N) \sum_i 1[x_i = x_0] \\ &\stackrel{p}{=} E(1[x = x_0] \cdot y) / P(x = x_0) = E(y|x = x_0). \end{aligned} \quad (35)$$

Folglich muss gelten

$$(1/N) \sum_i 1[x_i = x_0] y_i = E(1[x_0 = x_0] \cdot y) = E(y|x = x_0) \cdot P(x = x_0). \quad (36)$$

Analog funktioniert (34) für stetige x .

Da der Normierungsfaktor in g_N und f_N nicht N , sondern Nh^k ist, erhält man die asymptotische Verteilung von $r_N(x_0)$ durch Multiplikation von $r_N(x_0)$ mit $(Nh^k)^{0.5}$ und nicht \sqrt{N} . Da gilt

$$\begin{aligned} & (Nh^k)^{0.5} g_N(x_0) / f_N(x_0) - (Nh^k)^{0.5} g_N(x_0) / f(x_0) \\ &= (Nh^k)^{0.5} g_N(x_0) \{f_N(x_0)^{-1} - f(x_0)^{-1} = o_p(1)\}, \end{aligned} \quad (37)$$

kann aufgrund von $f_N(x_0) \stackrel{p}{=} f(x_0)$ die asymptotische Verteilung von $r_N(x_0)$ erhalten werden aus $(Nh^k)^{0.5} g_N(x_0) / f(x_0)$.

Die nonparametrische Kernel-Methode kann ebenso für die Schätzung anderweitiger Funktionen verwendet werden. Um beispielsweise die Varianz $V(y|x) = E(y^2|x) - \{E(y|x)\}^2$ zu berechnen, lässt sich $E(y^2|x)$ bestimmen, indem man y ersetzt durch y^2 . Des Weiteren lässt sich $\delta E(y|x_0) / \delta x = \delta r / \delta x$ schätzen durch Differenzieren von $r_N(x_0)$ nach x_0 . Ist $r(x_0) = x'_0 \beta$, dann ist $\delta r(x_0) / \delta x = \beta$. Daher reflektiert der

Schätzer für $\delta r(x_0)/\delta x$ den marginalen Effekt von x auf $r(\cdot)$ an der Stelle $x = x_0$. Um den Aussagegehalt solcher Effekte zu erhöhen, sollten diese tendenziell eher für „überglatte“ Regressionsfunktionen berechnet werden.

Was die Annahmen über die Kerne angeht, so kann auf (25) verwiesen werden. Es kann gezeigt werden, dass der Kernel-Schätzer konsistent ist, wenn $Nh^k \rightarrow \infty$ und $h \rightarrow 0$ bei gleichzeitig $N \rightarrow \infty$. Dann gilt $r_N(x_0) =^p r(x_0)$. Für uniforme Konsistenz, $\sup_x |r_N(x) - r(x)| = o_p(1)$, muss die maximale Abweichung von $r_N(x)$ von $r(x)$ gegen null konvergieren. Es kann gezeigt werden, dass unter bestimmten Umständen gilt: $\sup_x |r_N(x) - r(x)| =^{as} O(1/(N^{1-\epsilon}h^k)^{0.5})$.

Zur Bestimmung von Glättungsparameter und Kern haben sich in der Praxis datenbasierte Optimierungsmethoden durchgesetzt. Der „Leave-one-out“ Kernel-Schätzer für $r(x_j)$ lautet:

$$r_{Nj}(x_j) \equiv \sum_{i \neq j} K((x_i - x_j)/h) y_i / \sum_{i \neq j} K((x_i - x_j)/h). \quad (38)$$

h wird dann so gewählt, dass es das Kreuzvalidierungskriterium

$$(1/N) \sum_{j=1}^N \{y_j - r_{Nj}(x_j)\}^2 \cdot w(x_j) \quad (39)$$

minimiert, wobei $w(x_j)$ eine Gewichtungsfunktion darstellt. Diese soll die „Vorhersagefehler“ im Grenzbereich der Bandbreite verringern. Die konkrete Wahl der Funktion bleibt wieder dem Anwender überlassen. Im einfachsten Fall gilt $w(x_j) = 1$ für alle j . Dasjenige \hat{h} welches das Kreuzvalidierungskriterium minimiert, wurde von Härdle und Marron (1985) hergeleitet.

Denkbar wäre auch, dass es von Vorteil sein könnte, verschiedene Bandbreiten zu nutzen, da die Regressoren in der Regel unterschiedlich skaliert sind. Aufgrund des hohen Aufwands macht es jedoch mehr Sinn, alle Regressoren zu standardisieren und eine einheitliche Bandbreite zu verwenden. Auch bei der Kernel-Regression hat die Wahl der Bandbreite letztlich wieder größere Auswirkungen als die Wahl des Kerns. Allgemein sind Kerne mit begrenztem Träger zu bevorzugen. Bei diesen fällt die Verzerrung in den Grenzbereichen geringer aus. Punktweise Konfidenzbänder können ebenfalls ausgewiesen werden, selbst wenn Heteroskedastizität vorliegt⁵. Da eine mögliche Verzerrung in der Regressionsfunktion nicht unberücksichtigt bleiben sollte, wählt man in der Regel jedoch ein eher kleines h .

⁵ Siehe Härdle (1990).

2.3.3. Lokale polynomiale Regression

Als Verallgemeinerung der Kernel-Schätzung muss noch die lokale polynomiale Regression erwähnt werden. Der Schätzer $r(x)$ aus der Kernel-Regression stellt einen lokal konstanten Schätzer dar, da angenommen wird, dass er eine Konstante in der lokalen Umgebung von x_0 darstellt. Genauso kann man jedoch $r(x)$ linear in der Umgebung von x_0 darstellen, so dass in der Umgebung von x_0 gilt: $r(x) = a_0 + b_0(x - x_0)$. Der Schätzer der lokalen linearen Regression minimiert

$$\sum_{i=1}^N K\left(\frac{x_i - x_0}{h}\right) (y_i - a_0 - b_0(x_i - x_0))^2 \quad (40)$$

nach a_0 und b_0 , wobei $K(\cdot)$ eine Kernel-Gewichtungsfunktion ist. Dann gilt $\hat{a}_0 + \hat{b}_0(x - x_0)$ in der Umgebung von x_0 . Der Schätzer exakt bei x_0 ist dann $\hat{r}(x) = \hat{a}_0$ und \hat{b}_0 liefert einen Schätzer für die erste Ableitung $\hat{r}'(x_0)$. Allgemein minimiert ein lokaler polynomialer Schätzer vom Grad p

$$\sum_{i=1}^N K\left(\frac{x_i - x_0}{h}\right) \left(y_i - a_{0,0} - a_{0,1}(x_i - x_0) - \dots - a_{0,p} \frac{(x_i - x_0)^p}{p!}\right)^2, \quad (41)$$

was $\hat{m}^{(s)}(x_0) = \hat{a}_{0,s}$ liefert.

Der Hauptvorteil der lokalen polynomialen Regression besteht darin, dass die asymptotische Verzerrung nicht von $r'(x_0)$ abhängt und dass sie durch die lokale, lineare Anpassung an den Grenzbereichen der Bandbreite besser funktioniert. Möchte man eine Ableitung der Ordnung s schätzen, so wählt man p am besten als $p = s + 1$. Zur Schätzung einer ersten Ableitung wählt man daher einen quadratischen Schätzer.

2.4. Beitrag von Hoderlein und White (2009)

2.4.1. Die Idee

Unbeobachtbare, individuenspezifische Merkmale korrelieren in mikroökonomischen Anwendungen oft mit den Regressoren. Damit stellen sie eine unbeobachtbare Quelle von Endogenität dar und können zur Verzerrung der geschätzten Effekte der Regressoren auf die unabhängige Variable führen.

Verfügt man über Paneldaten, so bieten sich Lösungsmöglichkeiten wie in Abschnitt 2.3 erläutert an. Persistente Einflussfaktoren werden typischerweise mit skalar additiven Unbeobachtbaren modelliert (den Fixed Effects). Durch Bildung der ersten Differenzen kann für diese kontrolliert werden.

Hoderlein und White (2009) zeigen in ihrer Arbeit, dass bei vielen nichtseparablen Modellen eine bestimmte Form des Differenzierens angewendet werden kann. Als Spezialfall wird das Modell binärer Wahl mit additiv skalarem Fixed Effect betrachtet. Im Allgemeinen haben nichtseparable Modelle die folgende Form:

$$Y_t = \phi(X_t, Z_t, U_t, A), \quad t = 1, \dots, T. \quad (42)$$

$Y_t = Y_{it} \in \mathcal{Y}$ mit $i = 1, 2, \dots$, stellt einen beobachtbaren realwertigen Zufallsskalar dar. $(X_t, Z_t) = (X_{it}, Z_{it}) \in \mathcal{X} \times \mathcal{Z} \subseteq \mathbb{R}^{K+L}$ stellen beobachtbare realwertige Zufallsvektoren der Dimensionen K und L dar. Beide stellen fehlerfrei beobachtbare Einflussfaktoren für Y_t dar. $(U_t, A) = (U_{it}, A_i) \in \mathcal{U} \times \mathcal{A}$ stellen dagegen unbeobachtbare Einflussfaktoren dar, die jeweils zeitvariabel (U_t) bzw. zeitkonstant (A_i) sind. Beide können zählbar unendlich dimensional sein.

Der interessierende Effekt sei der von X_t auf Y_t . Für den Einfluss der anderen Variablen möchte man lediglich kontrollieren. Ob diese beobachtet werden oder nicht, spielt keine Rolle. Das konkrete Ausmaß der Effekte ist nicht von Interesse.

Der Einfachheit halber wird der Fall $T = 2$ angenommen. Es liegen also nur zwei Perioden vor. Es gilt damit: $\Delta Y := Y_2 - Y_1$, $X = (X'_1, X'_2)'$, $\Delta X := X_2 - X_1$, $Z = (Z'_1, Z'_2)'$ und $\Delta Z := Z_2 - Z_1$.

Das wichtigste Theorem der Arbeit lautet:

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &= \mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z]. \end{aligned} \quad (43)$$

Die erste Zeile enthält nur beobachtbare Einheiten. Die zweite Zeile entspricht der lokalen durchschnittlichen strukturellen Ableitung (Local Average Structural De-

rivative - LASD). Diese gilt für die Subpopulation, welche durch $\Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z$ charakterisiert wird. Dabei spielen X und Z unterschiedliche Rollen. Während nach den ersten Differenzen der X differenziert wird, wird für Z lediglich bedingt. Ist Z bzw. sind seine Komponenten zeitkonstant, dann ist $\Delta Z = 0$ automatisch erfüllt.

Es kann gezeigt werden, dass die LASD bereits unter schwachen Annahmen identifiziert ist. Notwendig sind lediglich bedingte Unabhängigkeit zwischen U_t und X_t , bedingt für A und Z sowie schwache Stationarität für den Fehler U . Betont sei dabei, dass Z beliebig korreliert sein kann mit A und U . Zwischen A und U selbst sowie zwischen A und X ist ebenfalls beliebige Korrelation möglich. Der Ansatz ermöglicht es damit in einer Vielzahl von Modellen den Einfluss potentiell endogener, persistenter, unbeobachtbarer Heterogenität zu berücksichtigen.

Zwei Schwächen sind jedoch zu bemängeln: Zum einen ist es nicht möglich, eine zeitlich verzögerte unabhängige Variable (zum Beispiel $y_{i,t-1}$) zu verwenden. Zum anderen sind selbst unter schwachen Annahmen lediglich Effekte für die Subpopulation zu identifizieren, für die gilt: $\Delta X = 0$ und $\Delta Z = 0$.

Besitzt der datengenerierende Prozess jedoch eine Indexstruktur, wie $\phi(X_t, Z_t, U_t, A) = \psi(X_t' \beta_0, Z_t' \gamma_0, U_t, A)$, dann reicht die Subpopulation für die Bestimmung relativer marginaler Effekte für die Gesamtstichprobe aus. Alternativ kann man die Annahmen verstärken und die Korrelation zwischen ΔX und A beschränken. Man erhält:

$$\begin{aligned} D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ = \mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) | X_1 = x, Z_1 = z]. \end{aligned} \quad (44)$$

Dies bedeutet, dass auf die LASD für die gesamte Population geschlossen werden kann, selbst wenn nur die Subpopulation mit $\Delta X = 0$ und $\Delta Z = 0$ betrachtet wird. Die zweite Schwäche des Ansatzes muss daher nicht als große Einschränkung betrachtet werden.

Dieser Ansatz kann auf das semiparametrische Paneldatenmodell binärer Wahl angewendet werden. Formal ausgedrückt gilt

$$Y_t = \mathbb{I}\{X_t' \beta_0 + Z_t' \gamma_0 + U_t + A > 0\}, \quad t = 1, \dots, T, \quad (45)$$

wobei alle Variablen wie gehabt sind. Es ist nun lediglich $\mathcal{U} \times \mathcal{A} \subseteq \mathbb{R}^2$, was bedeutet, dass A zur klassischen Notation der Fixed Effects korrespondiert. \mathbb{I} bezeichnet eine Indikatorfunktion und β und γ stellen unbekannte Koeffizienten dar.

Behält man $T = 2$ bei, so kann man zeigen, dass unter recht unrestrictiven Annahmen gilt:

$$\beta_0 \propto \mathbb{E}[\{D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1, \Delta Z = 0, Z] |_{\xi=0}\} b(X, Z)]. \quad (46)$$

Dabei stellt b eine nutzerspezifische Gewichtungsfunktion dar, die sich lediglich auf die Proportionalität auswirkt. Wie in semiparametrischen Indexmodellen üblich, ist leider nur die Identifikation von relativen Effekten möglich. Allgemein erhält man die Identifikationsergebnisse bei unbeobachtbarer Heterogenität durch Ersetzen von β_0 durch $\beta(A)$.

2.4.2. Angrenzende Literatur

Die Analyse nichtlinearer Paneldatenmodelle hat eine lange Tradition und geht zurück auf den bedingten ML-Ansatz von Rasch (1960, 1961). Parametrische Paneldatenmodelle wurden seitdem eingehend analysiert. Ein Überblick zu Modellen diskreter Wahl findet sich bei Arellano (2003).

Eng verbunden mit der Arbeit von Hoderlein und White (2009) ist der Ansatz von Manski (1987). In diesem wird die semiparametrische Schätzung nicht dynamischer Paneldatenmodelle binärer Wahl über eine Median-Restriktion betrachtet. Chamberlain (2010) behandelt die Identifikation im dynamischen Paneldatenmodell binärer Wahl und klärt warum die Annahme der logistischen Verteilung zur Identifikation von β_0 in der Regel benötigt wird. Nichtlineare Fixed Effects Modelle finden sich außerdem bei Hausmann, Hall und Griliches (1984) für ein Panel mit Zähldaten, bei Honoré (1992) für ein Panel mit zensierten Daten und bei Kyriazidou (1997) für ein Panel mit einem Modell zur Stichprobenselektion.

Wie in all diesen anderen Arbeiten auch, wird im Ansatz von Hoderlein und White (2009) eine fixe Zahl von Zeitperioden angenommen. Das besondere ist, dass nur zwei Perioden benötigt werden. Dies ist ein entscheidender Unterschied zu anderen Ansätzen, die oft nicht nur mehr Perioden benötigen, sondern auch nur Aussagen für sehr begrenzte Subgruppen ermöglichen.⁶ In einigen Ansätzen wird sogar verlangt, dass die Zahl der Zeitperioden gegen unendlich geht.⁷ Eine interessante Möglichkeit zum Umgang mit der in zahlreichen dynamischen Modellen mit fixem T auftretenden Verzerrung, zeigen Honoré und Tamer (2006). Ein allgemeiner Überblick kann Chamberlain (1984) sowie Arellano und Honoré (2001) entnommen werden.

⁶ Siehe beispielsweise Honoré und Kyriazidou (2000).

⁷ Siehe Arellano und Hahn (2007) sowie Hahn und Newey (2004).

Alle erwähnten Ansätze behandeln ein bestimmtes semiparametrisches Modell wie das dynamische Modell binärer Wahl. Eine von der Grundidee dem Ansatz von Hoderlein und White (2009) ähnliche Arbeit, stellt die Arbeit von Chernozhukov et al. (2009) dar. In dieser wird der Fall diskreter Variablen betrachtet. Ebenfalls ähnlich sind die Ansätze von Graham und Powell (2009) sowie von Arellano und Bonhomme (2009). In beiden liegt der Fokus auf einer linear heterogenen Population und nicht auf einer nichtseparablen Struktur. Beide benötigen mindestens so viele Zeitperioden wie Regressoren. Das von Altonji und Matzkin (2005) behandelte Modell ist ebenfalls interessant, liegt jedoch eher dem Random Effects Modell nahe. Im Modell von Bester und Hansen (2009) wird schließlich ein Fixed Effects Modell behandelt, in dem die Regressoren schwach separabel von den korrelierten Unbeobachtbaren sind.

2.4.3. Identifikation marginaler Effekte in nichtseparablen Funktionen

Begleitet von einigen Anmerkungen und Beispielen, werden im Folgenden zunächst die wichtigsten Annahmen und Schreibweisen eingeführt. Im Zwei-Perioden-Fall lässt sich bilden:

$$\begin{aligned} Y_1 &= \phi(X_1, Z_1, U_1, A) \\ Y_2 &= \phi(X_2, Z_2, U_2, A). \end{aligned} \tag{47}$$

Dabei sei Y_t eine skalare Zufallsvariable und X_t und Z_t seien Zufallsvektoren finiter Dimension K und L . U_t und A seien Zufallsvektoren für die unbeobachtbaren Variablen mit beliebiger, jedoch abzählbar unendlicher Dimension. Angenommen sei weiterhin eine ausreichend reguläre bedingte kumulative Verteilungsfunktion.⁸ Damit ist abgesichert, dass $F_{A|\Delta X, X_1, \Delta Z, Z_1}(a|\xi, x, 0, z)$ eine Dichte der Form

$$f(a|\xi, x, 0, z)\mu(da|x, 0, z) \tag{48}$$

besitzt. f wird als Radon-Nikodým Ableitung verstanden.⁹ Ist A stetig, dann entspricht diese einer bedingten Dichtefunktion. A ist dann absolut stetig im Hinblick auf das Lebesgue-Maß μ .¹⁰ Ist A jedoch diskret, dann entspricht sie einer bedingten Wahrscheinlichkeit von A . μ ist dann Zählmaß.¹¹

⁸ Die Funktion ist dann lokal integrierbar und es gilt $F(\infty) = 1$.

⁹ Der Satz von Radon-Nikodým verallgemeinert die Ableitung einer Funktion auf Maße. Unter einem Maß wird in der Maßtheorie eine Zuordnung von reellen oder komplexen Zahlen zu einem Teilmengensystem über einer Grundmenge verstanden. Näheres siehe Elstrodt (2005), S. 279 ff.

¹⁰ Ein Maß, welches geometrischen Objekten ihren Inhalt (beispielsweise die Länge oder Fläche) zuordnet. Näheres siehe Elstrodt (2005), S. 279 ff.

¹¹ Eine Funktion, die jeder Menge die Anzahl ihrer Elemente zuordnet. Näheres siehe Elstrodt (2005), S. 29.

Unter dem Ausdruck, dass die Bedingungen „fast überall- μ “ gelten, wird im Folgenden verstanden, dass diese im Hinblick auf $\mu(\cdot|x, 0, z)$ gelten. Außerdem sei $\nu(u, a|x, z)$, das Produktmaß, definiert durch $F(u|a, 0, z) \times \mu(a|x, 0, z)$.¹² Funktionen die „ ν -integrierbar“ sind, werden verstanden als integrierbar nach $\nu(\cdot, \cdot|x, z)$. $\|\xi\| := [\xi'\xi]^{1/2}$ stellt die Euklidische Norm dar und die Umgebung sei $\mathcal{N}_\epsilon = \{\xi : \|\xi\| < \epsilon\}$. Schließlich stehe $|M| := \max_{ij} |M_{i,j}|$ für jede Matrix $M := [M_{i,j}]$.

Nun aber zu den Annahmen, die es ermöglichen, die LASD zu identifizieren:

Annahme 1:

Es sei (Ω, \mathcal{F}, P) ein vollständiger Wahrscheinlichkeitsraum auf dem folgende Zufallsvektoren definiert sind: $A : \Omega \rightarrow \mathcal{A}, \mathcal{A} \subseteq \mathbb{R}^\infty$ und $(Y_t, X_t, Z_t, U_t) : \Omega \rightarrow \mathcal{Y} \times \mathcal{X} \times \mathcal{Z} \times \mathcal{U}, \mathcal{Y} \subseteq \mathbb{R}, \mathcal{X} \subseteq \mathbb{R}^K, \mathcal{Z} \subseteq \mathbb{R}^L, \mathcal{U} \subseteq \mathcal{R}^\infty, t = 1, 2$, mit K und L endlichen ganzen Zahlen. Für $t = 1, 2$ gilt erstens $\mathbb{E}(Y_t) < \infty$, zweitens $Y_t = \phi(X_t, Z_t, U_t, A)$, wobei $\phi : \mathcal{X} \times \mathcal{Z} \times \mathcal{U} \times \mathcal{A} \rightarrow \mathcal{Y}$ eine Borel-messbare Funktion ist und drittens die Realisationen von (Y_t, X_t, Z_t) beobachtbar sind, während die von (U_t, A) unbeobachtbar sind.¹³

Diese Annahme spezifiziert formal den datengenerierenden Prozess, der bereits im Vorangegangenen diskutiert wurde. Dadurch, dass ϕ zeitkonstant ist, werden unbeschränkte Zeittrends ausgeschlossen. Es gibt jedoch bestimmte Möglichkeiten, diese dennoch zu berücksichtigen. Dies wird an späterer Stelle erläutert.

¹² Entspricht einem Maß auf dem Produkt zweier Maßräume. Näheres siehe Elstrodt (2005), S. 163 ff.

¹³ Borel-Messbarkeit siehe Elstrodt (2005), S. 86 ff. sowie Bierens (2005), S. 37.

Annahme 2:

Es existiert ein $\epsilon > 0$, so dass

$$U_t \perp (\mathbb{I}\{\|\Delta X\| < \epsilon\} \Delta X, X_1) | A, \mathbb{I}\{\|\Delta Z\| = 0\} \Delta Z, Z_1 \quad t = 1, 2. \quad (49)$$

Diese Annahme spezifiziert die Art der Exogenität von X . Bedingt für A , Z_1 und $\Delta Z = 0$, ist U_1 unabhängig von X_1 und ΔX (für U_2 analog). Für die Differenzen ΔX muss die Unabhängigkeitsbedingung jedoch lediglich für geringe Änderungsbeiträge gelten, während Differenzen von größerem Ausmaß außer Acht bleiben. Des Weiteren können Z_1 und ΔZ beliebig korreliert sein mit U_t , wobei der U_t -Prozess prinzipiell Autokorrelation aufweisen kann.

Diese Annahme kann ökonomisch interpretiert werden. Man nehme an, man habe einen Datensatz mit individuenspezifischen Informationen für verschiedene Gruppen von Individuen. Erfasst sei die Nachfrage nach einem bestimmten Gut, die Höhe der Haushaltseinkommen sowie weitere Haushaltscharakteristika. Zur Vereinfachung seien letztere im Zeitverlauf unveränderlich. Außerdem sei angenommen, dass der Effekt der Haushaltscharakteristika nicht von Interesse ist (diese spielen die Rolle von Z). Für den Einfluss der Haushaltscharakteristika auf das Einkommen X möchte man lediglich kontrollieren. Der Effekt des Einkommens auf die Nachfrage soll möglichst unverzerrt wiedergegeben werden. Nun sei angenommen, A und Z_t seien diskret und man unterteile die Grundgesamtheit anhand der individuellen Kriterien (z, a) , welche hier als „Zelle“ bezeichnet seien. Dann verlangt Annahme 2, dass die Fehler (z.B. die transitorischen Schocks) U_1 und U_2 marginal aber nicht zwingend gemeinsam unabhängig vom Einkommen X_1 sind und dessen Änderungen ΔX innerhalb jeder Zelle.

Hätte man nun Daten über das Geschlecht und die Branche in der ein Individuum arbeitet in Z enthalten und in A steckten unbeobachtete Fähigkeiten, dann würde dies bedeuten, dass U_t unabhängig verteilt sein müsste vom Einkommen und von kleinen Einkommensänderungen und zwar für zum Beispiel alle weiblichen Arbeiterinnen mit bestimmten Fähigkeiten, die in der metallverarbeitenden Industrie arbeiten. Würden die (z, a) Werte ignoriert, so könnten Einkommen und transitorische Schocks U_t jedoch abhängig sein.

Interessiert man sich für die Effekte sämtlicher Regressoren, so gibt es leider keine Z für die man bedingen kann. Annahme 2 rückt dann der strikten Exogenitätsbedingung aus den Lehrbüchern für lineare Paneldatenmodelle näher. Der einzige große Unterschied ist dann, dass Korrelation zwischen A und U erlaubt ist und dass A und

U beliebig große Dimensionen annehmen können. Diese bereits schwache Bedingung kann noch weiter abgeschwächt werden, indem zusätzliche bedingende Instrumente wie zum Beispiel Vergangenheitswerte von X herangezogen werden.

Annahme 3:

U_t ist bedingt stationär: $F_{U_1|A, \mathbb{I}\{\|\Delta Z\|=0\}\Delta Z, Z_1} = F_{U_2|A, \mathbb{I}\{\|\Delta Z\|=0\}\Delta Z, Z_2}$.

Diese Annahme entspricht einer schwachen Stationaritätsannahme für die unbeobachtbaren Einflussfaktoren. Die Hauptaussage ist, dass die bedingte Verteilung von U_t zeitkonstant ist. Wie im Folgenden deutlich wird, schließt diese Annahme die Berücksichtigung zeitlich verzögerter Werte der abhängigen Variablen als Regressoren aus.

Annahme 4:

\mathcal{X} ist eine offene konvexe Menge und für jedes $(z, u, a) \in \mathcal{Z} \times \mathcal{U} \times \mathcal{A}$ ist $\phi(\cdot, z, u, a)$ zweifach stetig differenzierbar auf \mathcal{X} . Außerdem sei $\mathbb{E}[D_x \phi(X_1, Z_1, U_1, A)] < \infty$ und $\mathbb{E}[D_{xx} \phi(X_1, Z_1, U_1, A)] < \infty$.

In dieser Annahme wird spezifiziert, dass die Funktion ϕ differenzierbar ist und zwar nach denjenigen Variablen die von Interesse sind. Die Bedingungen sichern ab, dass die benötigten Erwartungswerte definiert sind.

Annahme 5:

Für jedes $(x, z) \in \mathcal{X} \times \mathcal{Z}$ gibt es ein σ - endliches Maß $\mu(\cdot|x, 0, z)$, welches absolut stetig ist im Hinblick auf $F(\cdot|\xi, x, 0, z)$ für alle $\xi \in \mathcal{N}_\epsilon$, so dass eine Radon-Nikodým Dichte f existiert und somit für jedes $(a, \xi) \in \mathcal{A} \times \mathcal{N}_\epsilon$ gilt, dass $F(da|\xi, x, 0, z) = f(a|\xi, x, 0, z)\mu(da|x, 0, z)$.

Damit darf die bedingte Wahrscheinlichkeit von A und Z abhängen von realisierten Werten von ΔX und X_1 . Es ist jedoch nicht erlaubt, dass die möglichen Werte für A und Z von den realisierten Werten ΔX und X_1 abhängen.

Annahme 6:

Für jedes $(x, z) \in \mathcal{X} \times \mathcal{Z}$, existiert $D_\xi f(a|\xi, x, 0, z)$ „fast überall - μ “ für alle $\xi \in \mathcal{N}_\epsilon$.

Diese Annahme fordert, dass $f(a|\xi, x, z)$ stetig differenzierbar ist für $\xi = 0$. Die Differenzierbarkeit in Kombination mit der Tatsache, dass man für X eine Umgebung von null Änderungen betrachtet, bedeutet dass X stetig verteilt sein muss. Diskrete Zufallsvariablen X sind damit ausgeschlossen. Angemerkt werden muss jedoch, dass nicht bedingt werden muss für die Kovariate Z und dass keine Restriktion für die Korrelation von Z mit den unbeobachtbaren Variablen aufgestellt wird.

Annahme 7:

Für jedes $(x, z) \in \mathcal{X} \times \mathcal{Z}$, gibt es eine v - integrierbare dominante Funktion $(u, a) \rightarrow D(u, a|x, z)$ so dass

$$\begin{aligned} \sup_{\xi \in \mathcal{N}_\epsilon} |D_{xx}\phi(x + \xi, z, u, a)f(a|\xi, x, 0, z)| &\leq D(u, a|x, z) \\ \sup_{\xi \in \mathcal{N}_\epsilon} |D_x\phi(x + \xi, z, u, a)'D_\xi f(a|\xi, x, 0, z)| &\leq D(u, a|x, z). \end{aligned} \tag{50}$$

Diese Annahme kann ebenso wie Annahme 5 als Regularitätsbedingung aufgefasst werden und ermöglicht die entsprechenden Auf- und Ableitungen.

Annahme 8:

$$A \perp (\mathbb{I}\{\|\Delta X\| < \epsilon\}\Delta X, \mathbb{I}\{\|\Delta Z\| = 0\}\Delta Z) | X_1, Z_1.$$

Diese letzte Annahme beschränkt die Korrelation zwischen den Inkrementen der X_t und Z_t Prozesse und A bedingt für X_1, Z_1 . Es sei lediglich erwähnt, dass diese Einschränkung nicht generell gelten muss und für das Hauptresultat der Arbeit nicht benötigt wird. Erwähnt werden muss jedoch, dass sie auch mit korrelierten A erfüllt wird, wenn der Einfluss von A additiv separabel ist von anderen Einflussfaktoren von X , wie zum Beispiel bei $X_t = \psi(X_{t-1}, Z_t, U_t) + \lambda(A)$.

Die eingeführten Annahmen erlauben nun die Identifikation der LASD. Das Hauptresultat lautet folgendermaßen:

Theorem 1:¹⁴

Sind die Annahmen 1 bis 7 erfüllt, so gilt

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &= \mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z] \end{aligned} \quad (51)$$

mit einer Wahrscheinlichkeit von eins. Ist zusätzlich Annahme 8 erfüllt, dann gilt:

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &= \mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) | X_1 = x, Z_1 = z] \end{aligned} \quad (52)$$

mit einer Wahrscheinlichkeit von eins.

Das Hauptresultat besagt, dass bestimmte bedingte Durchschnitte der Ableitungen identifiziert sind. Die jeweils obere Zeile enthält nur beobachtbare Variablen: Sie stellt die Ableitung der nichtparametrischen Regression von ΔY auf $\Delta X, X_1, \Delta Z, Z_1$ nach den ersten Argumenten (also ΔX) berechnet für beliebige Werte von $X_1 = x, Z_1 = z$ und für $\Delta X = 0, \Delta Z = 0$ dar. Die untere Zeile entspricht in (44) exakt der bereits eingeführten LASD.

Um die ökonomische Relevanz des Theorems zu verdeutlichen, sei wieder das Beispiel der Konsumentennachfrage angeführt. Zusätzlich seien nun noch die zeitlich variablen Bildungsjahre der Personen verfügbar. Laut dem ersten Teil von Theorem 1 wäre dann der durchschnittliche marginale Effekt des Einkommens auf die Nahrungsmittelnachfrage bestimmt und zwar beispielsweise für die Gruppe weiblicher Arbeiterinnen in der metallverarbeitenden Industrie mit zehn Bildungsjahren und einem Einkommen von mehr als 50.000 Euro. Vorausgesetzt wird jedoch, dass sich Einkommen und Bildungsjahre nicht über die Perioden hinweg verändern. Für einzelne, individuelle Frauen lassen sich keine marginalen Effekte bestimmen. Unberücksichtigte persistente Faktoren wie zum Beispiel Präferenzen stellen jedoch kein Problem dar. Diese dürfen beliebig mit dem Einkommen, der Beschäftigung oder den Bildungsjahren korreliert sein.

Im zweiten Teil des Theorems kommt die stärkere Annahme 8 hinzu. Ist diese zusätzlich erfüllt, so ist beispielsweise der durchschnittliche marginale Effekt für alle weiblichen Arbeiterinnen in der metallverarbeitenden Industrie mit 10 Bildungsjahren und einem Einkommen von mehr als 50.000 Euro Einkommen bestimmbar. Ob

¹⁴ Für die Herleitung sei auf den Anhang in Hoderlein und White (2009) verwiesen.

sich das Einkommen ändert oder die Bildungsjahre im Zeitverlauf zunehmen, spielt keine Rolle mehr.

Ein möglicher Kritikpunkt am obigen Theorem ist, dass es lediglich für eine kleine Subpopulation gültig ist. Nimmt man jedoch an, dass die zweiten Ableitungen von ϕ und $D_\xi f(a|\xi, x, 0, z)$ uniform beschränkt sind, dann folgt daraus, dass die Verzerrung am meisten von der Differenz ξ abhängt. Ist ξ gering (bzw. annähernd null), so ist zu erwarten, dass nur eine geringe Verzerrung vorliegt.

Ausgehend von Theorem 1 sind einige interessante Erweiterungen möglich. Diese werden im Folgenden betrachtet.

Erweiterung 1: Zeitverzögerte Variablen und Trends

Erstens ist es prinzipiell doch möglich, zeitlich verzögerte Regressoren zu berücksichtigen. Betrachtet werde der Drei-Perioden-Fall:

$$\begin{aligned}\mathcal{Y}_1 &= \phi(\mathcal{X}_1, \mathcal{X}_0, \mathcal{U}_1, A) \\ \mathcal{Y}_2 &= \phi(\mathcal{X}_2, \mathcal{X}_1, \mathcal{U}_2, A) \\ \mathcal{Y}_3 &= \phi(\mathcal{X}_3, \mathcal{X}_2, \mathcal{U}_3, A).\end{aligned}\tag{53}$$

Die bisherige Identifikationsstruktur funktioniert genauso, wenn man Zeitdifferenzen in der abhängigen Variable einsetzt, die weiter reichen als die Zeitverzögerungen der abhängigen Variable. Schreibt man um: $X_2 = (\mathcal{X}'_3, \mathcal{X}'_2)'$, $X_1 = (\mathcal{X}'_1, \mathcal{X}'_0)'$, $Y_2 = \mathcal{Y}_3$, $Y_1 = \mathcal{Y}_1$, $U_2 = \mathcal{U}_3$ und $U_1 = \mathcal{U}_1$, dann passt dies genau zum Bisherigen. Man verwendet also $\Delta\tilde{Y} = Y_2 - Y_1 = \mathcal{Y}_3 - \mathcal{Y}_1$ und $\Delta X = (\mathcal{X}'_3 - \mathcal{X}'_1, \mathcal{X}'_2 - \mathcal{X}'_0)'$. Unter lediglich geringfügigen Modifikationen kann damit das bisherige Resultat aufrecht erhalten bleiben.

Zweitens können auch die bisher ausgeschlossenen unbeschränkten Zeittrends eingeschlossen werden. Dazu muss angenommen werden:

$$Y_t = \phi(t, X_t, Z_t, U_t, A) = \phi_0(X_t, Z_t, U_t, A) + \phi_1(t, U_t, Z_t).\tag{54}$$

X_t , A und t sind additiv separabel. Dies stellt eine Verallgemeinerung des allgemein verwendeten additiven Zeittrends dar, ist jedoch in so weit restriktiv, als dass die marginalen Effekte $D_x\phi$ nicht von t abhängen dürfen.

Zu diesen Betrachtungen ähnliche Analysen erlauben nicht nur die Identifikation der durchschnittlichen marginalen Effekte bei Vorliegen von generalisierten Fixed Effects, sondern auch die Identifikation marginaler kausaler Effekte bezüglich verschiedenster interessierender Aspekte der bedingten Antwort-Verteilung. Dies ist vor allem im Rahmen der Politikevaluation von Interesse.

Das obige Resultat ist zudem auf bekannte, differenzierbare Transformationen von Y anwendbar. Zum Beispiel sei $g(y) = F_Y(y)$, wobei F_Y die bedingte Verteilungsfunktion von Y darstellt. Dann folgt aus der Argumentation in Theorem 1:

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta g(Y) | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ & \mathbb{E}[D_x(g(\phi(X_1, Z_1, U_1, A)) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z] \\ & \mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) g'(Y) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z]. \end{aligned} \quad (55)$$

Damit sind gewichtete Durchschnitte der Form $\mathbb{E}[D_x \phi f_Y(Y) | \cdot]$ identifiziert. Genauso aber auch Gewichtungsschemata, die es erlauben, dass man eine Untergruppe von Y allein betrachtet. Insbesondere in der Politikevaluation ist dies von Interesse, da hier in der Regel nicht bloß die durchschnittlichen marginalen Effekte von Interesse sind, sondern die marginalen Effekte an bestimmten Punkten der Y -Verteilung.

Tatsächlich muss g nicht einmal differenzierbar sein. Betrachtet werde die bedingte Verteilungsfunktion, welche man durch $g_y(\phi) = \mathbb{I}\{\phi \leq y\}$ erhält. Diese ist nicht differenzierbar. Das Ergebnis lässt sich jedoch genauso herleiten wie im folgenden Abschnitt bei Vorhandensein einer binären Variable. Man erhält:

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta g_y(Y) | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ & \mathbb{E}[D_x \psi_y(X_1, Z_1, A) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z], \end{aligned} \quad (56)$$

wobei

$$\begin{aligned} \psi_y(x, z, a) &= \int \mathbb{I}\{\phi(x, z, u, a) \leq y\} F_{U|A\mathbb{I}\{|\Delta Z|=0\}\Delta Z, Z_1}(du | a, 0, z) \\ &= \mathbb{P}[Y_1 \leq y | X_1 = x, Z_1 = z; \Delta Z = 0, A = a] \end{aligned} \quad (57)$$

annahmegemäß nach x differenzierbar ist. Der Ausdruck ermöglicht wieder eine LASD-Interpretation. Die interessierende strukturelle Ableitung ist $D_x \psi_y(x, z, a)$. Der Durchschnittseffekt für die gegebenen Kovariate entspricht exakt dem erhaltenen Effekt.

Erweiterung 2: Bedingte Endogenität von X

Selbst wenn man die interessierenden marginalen Effekte vom Einfluss individuenspezifischer persistenter Heterogenität isoliert hat, kann theoretisch immernoch Abhängigkeit zwischen U_t und X_t bestehen. Es könnte sein, dass Annahme 2 nicht erfüllt ist, wenn man lediglich für die vorab spezifizierten Variablen bedingt. Ist X_t beispielsweise eine binäre Entscheidungsvariable und stellt U_t ganz neue Informationen für den Entscheidungsträger dar, dann könnte ein Endogenitätsproblem auftreten.

Lösen lässt sich dies, indem man zusätzliche Kontrollvariablen einführt. Man nehme an, X_t sei strukturell generiert als

$$X_t = \mathcal{X}(W_t, V_t, A), t = 1, 2, \quad (58)$$

wobei W_t und V_t beobachtbare und unbeobachtbare Faktoren für X_t darstellen und \mathcal{X} eine unbekannte messbare Funktion darstellt. \mathcal{X} sei ausreichend invertierbar in V_t . Der Vorteil von Paneldaten liegt normalerweise darin, dass es einfache Kandidaten für W_t gibt: zum Beispiel vergangene X_t 's. Liegt Endogenität vor, weil gleiche Informationen in die Entscheidungen für X_t und Y_t einfließen, dann kann angenommen werden, dass vergangene Entscheidungsvariablen auch nur vergangene Informationen enthalten und von künftigen Informationen unabhängig sind.

In einer passenden Struktur kann V_t jedoch als Kontrollvariable wiederhergestellt werden. Nimmt man beispielsweise $X_t = \mathcal{X}_0(W_t) + \mathcal{X}_1(W_t, A)V_t$ mit $V_t \perp (W_t, A)$ an und führt die Normalisierungen $\mathbb{E}(V_t) = 0, \text{Var}(V_t) = 1$ ein, dann ermöglicht dies, nach V_t aufzulösen, da $V_t = \text{Var}(X_t|W_t)^{-1/2}[X_t - \mathbb{E}(X_t|W_t)]$, und $\mathbb{E}(X_t|W_t)$ und $\text{Var}(X_t|W_t)$ einfach geschätzt werden können.

Unter Exogenitätsbedingungen für W_t , analog zu solchen die die Validität von Standard-Instrumenten-Variablen absichern, können nun V_1 und $\Delta V := V_2 - V_1$ als Kontrollvariablen verwendet werden. Insbesondere kann angenommen werden, dass

$$(U_t, \Delta V, V_1) \perp (W_1, W_2) | A, \mathbb{I}\{||\Delta Z|| = 0\} \Delta Z, Z_1 \quad t = 1, 2. \quad (59)$$

Dies kann erweitert werden zu¹⁵

$$U_t \perp (\mathbb{I}\{\|\Delta X\| < \epsilon\} \Delta X, X_1) | A, \mathbb{I}\{\|\Delta Z\| = 0\} \Delta Z, Z_1, \Delta V, V_1 \quad t = 1, 2 \quad (60)$$

und stellt eine Abwandlung von Annahme 2 dar, in der Z_1 um die Kontrollvariablen ΔV und V_1 angereichert wird.

Erweiterung 3: Chamberlain's Unmöglichkeitstheorem

Wie bereits erwähnt, deckt Theorem 1 den Fall zeitlich verzögerter abhängiger Variablen eigentlich nicht ab. Betrachtet sei folgendes Gleichungssystem:

$$\begin{aligned} Y_1 &= \phi(Y_0, X_1, U_1, A) \\ Y_2 &= \phi(Y_1, X_2, U_2, A). \end{aligned} \quad (61)$$

Auf den ersten Blick erscheint es so als ob die Möglichkeit bestünde, lokale durchschnittliche strukturelle Effekte zu identifizieren, indem man Y_{t-1} als einen zusätzlichen interessierenden Einflussfaktor betrachtet. Dabei verstößt man jedoch gegen Annahme 2. Eine Alternative wäre es, Y_{t-1} als zusätzliche bedingende Variable zu verwenden (zum Beispiel $Z_t = Y_{t-1}$). Doch auch dies ist nicht möglich. Man verstößt gegen die sonst eher hintergründige Annahme 3. Insbesondere ist es nicht möglich, dass

$$U_2 | A, Y_1, Y_0 \sim U_1 | A, Y_1, Y_0, \quad (62)$$

denn obwohl U_2 plausiblerweise als unabhängig von Y_1 und Y_0 angenommen wird (zum Beispiel wenn es keinerlei serielle Korrelation zwischen den U_t gibt), wird die Annahme nicht erfüllt, weil U_1 dabei hilft, Y_1 zu bestimmen.

Erweiterung 4: Schätzung

Grundsätzlich ist die Schätzung der interessierenden Größen relativ unkompliziert. Standardmäßige nonparametrische Regressionsverfahren können angewendet werden. Liegt das Interesse auf bestimmten partiellen Ableitungen oder einer Reihe von partiellen Ableitungen für bestimmte Werte der bedingenden Variablen (zum Beispiel deren Durchschnitte), so ist die Kernel-Regression besonders zu empfehlen.

¹⁵ Siehe Dawid (1979).

Die lokale polynomiale Regression wie sie beispielsweise von Cleveland (1979) vorgeschlagen wurde, stellt jedoch eine brauchbare Alternative dar. Die Methoden sind relativ leicht nachvollziehbar und liefern selbst unter schwachen Annahmen Bedingungen konsistente und asymptotisch normale Schätzer für die Ableitungen.¹⁶ Außerdem lässt sich der empirisch bedeutsame Fall gemischter Daten (einige Variablen sind stetig, andere diskret) behandeln und die Probleme an den Grenzbereichen, wie sie bei der Kernel-Regression auftauchen, können vermieden werden.

Verfügt man über ein Panel der Individuen $i = 1, \dots, n$, von dem der Einfachheit halber angenommen wird, dass die Beobachtungen unabhängig und identisch verteilt (iid) sind, dann erhält man die Parameterschätzer aus der lokalen Polynom-Regression als:

$$\hat{\theta}_n(w) = \arg \min_{\theta \in \Theta} \sum_{i=1}^n [\Delta Y_i - g_p(W_i, \theta)]^2 K_{hn}(W_i - w), \quad (63)$$

wobei $g_p(W, \theta)$ ein Polynom vom Grad p definiert in $W := (\Delta X', X_1', \Delta Z', Z_1')'$ mit Parametern θ ; K_{hn} ist ein multivariater Kern mit passend gewählter Bandbreite h_n , zum Beispiel dem Produktkern

$$K_{hn}(W_i - w) = h_n^{-d} \prod_{l=1}^d \kappa\left(\frac{W_{i,l} - w_l}{h_n}\right), \quad (64)$$

wobei $d := 2K + 2L$, κ ein univariater Kern ist und W_i die Elemente $W_{i,l}$ besitzt und $w_0 := (\xi, x, 0, z)$, $\xi \in \mathcal{N}_\epsilon$ die interessierenden Kovariate definiert. Es handelt sich damit um reine Standardverfahren.

Hat man den Schätzer $\hat{\theta}_n$, so bestimmt man den Effekt, der von Interesse ist:

$$\begin{aligned} \delta^*(x, z) &:= D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &\mathbb{E}[D_x \phi(X_1, Z_1, U_1, A) | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z], \end{aligned} \quad (65)$$

da

$$\hat{\delta}_n(x, z) := D_\xi g_p(w_0, \hat{\theta}_n(w_0)) |_{\xi} = 0. \quad (66)$$

Unter schwachen Annahmen¹⁷ resultiert

$$\sqrt{nh^{d+2}}(\hat{\delta}_n(x, z) - \delta^*(x, z) - h^p \mathcal{B}(x, z)) \xrightarrow{d} \mathcal{N}(0, \Sigma(x, z)), \quad (67)$$

¹⁶ Siehe Fan (1992), Ruppert und Wand (1994), Fan und Gijbels (1996).

¹⁷ Siehe zum Beispiel Fan und Gijbels (1996) oder Hoderlein (2005).

wobei $\mathcal{B}(x, z)$ und $\Sigma(x, z)$ die asymptotische Verzerrung und die Kovarianzmatrix darstellen.

Erweiterung 5: Durchschnittliche Effekte

Von Interesse können auch die Durchschnitte der bedingten Effekte sein. Insbesondere die Annahmen, die zur Identifikationsgleichung (50, 51) führen, erlauben es Durchschnittseffekte der Form

$$\mathbb{E}_F[D_x\phi(X, Z, U_1, A)] := \int \mathbb{E}[D_x\phi(x, z, U_1, A)|X = x, Z = z]F(dx, dz) \quad (68)$$

zu identifizieren, wobei F eine Dichte darstellt, die für den Anwender von Interesse ist. Insbesondere liegt vor:

$$\begin{aligned} \delta_0^* &:= \int D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} F(dx, dz) \\ &=: \mathbb{E}_F[D_x\phi(X, Z, U_1, A)], \end{aligned} \quad (69)$$

was dem Gesamt-Durchschnittseffekt über die Grundgesamtheit entspricht, für welche $\Delta X = 0$ und $\Delta Z = 0$ gilt. Diese Größe entspricht einem partiellen Durchschnitt und kann daher geschätzt werden über den Durchschnitt

$$\widehat{\delta}_{0,n} := n^{-1} \sum_{i=1}^n \widehat{\delta}_n(X_i, Z_i), \quad (70)$$

wobei $\{X_i, Z_i\}$ iid ist mit gemeinsamer Verteilung, für die Newey (1994) Bedingungen liefert, unter denen

$$n^{1/2} h_n^\alpha (\widehat{\delta}_{0,n} - \delta_0^*) \xrightarrow{d} \mathcal{N}(0, \sum_0), \quad (71)$$

wobei α eine Konstante ist, die von den Dimensionen von X und Z abhängt. \sum_0 ist eine Kovarianzmatrix, deren spezifische Form von der Wahl von g_p abhängt. Im Anhang der Arbeit von Hoderlein und White (2009) wird der Schätzer für \sum_0 für das lokale lineare Polynom g_1 hergeleitet. Die Ableitung für $p > 1$ geht daraus einfach hervor.

2.4.4. Endogenes Modell binärer Wahl mit Heterogenität

Nun sei explizit der Fall mit einer binären abhängigen Variable Y_t betrachtet. Außerdem sei X_t möglicherweise mit U_t und A korreliert. Um aussagekräftige Ergebnisse zu erhalten, müssen die vorangegangenen Annahmen erweitert bzw. modifiziert werden:

Annahme 9:

Annahme 1 gilt mit

$$\phi(X_t, Z_t, U_t, A) = \mathbb{I}\{X_t' \beta(A) + Z_t' \gamma_0 + U_t + \alpha(A) > 0\}, \quad (72)$$

wobei $\alpha : \mathcal{A} \rightarrow \mathbb{R}$ und $\beta : \mathcal{A} \rightarrow \mathbb{R}^k$ unbekannte messbare Funktionen sind, γ_0 ist ein unbekannter endlicher $L \times 1$ Vektor und U_t ist ein Zufallsskalar.

Dies reduziert sich auf das Lehrbuch-Modell binärer Wahl mit Fixed Effects, wenn für alle $a \in \mathcal{A}$, $\beta(a) = \beta_0$, wobei β_0 ein unbekannter Vektor ist.

In dieser Annahme wird wieder der datengenerierende Prozess formal spezifiziert. Betrachtet wird eine latente Variable, die von einer linearen Struktur bestimmt wird. Die Koeffizienten variieren über die Population hinweg als eine Funktion der persistenten unbeobachtbaren Einflussfaktoren A (A stellt zum Beispiel wieder individuelle Präferenzen dar). U_t fließt auf eine additiv separable Weise ein.

Annahme 10:

Annahme 3 gilt und für jedes $(a, z) \in \mathcal{A} \times \mathcal{Z}$, ist $u \rightarrow F_{U|A, \mathbb{I}\{\|\Delta Z\|=0\}\Delta Z, Z_1}(u|a, 0, z)$ zweifach stetig differenzierbar nach u für alle $u \in \mathcal{U}$ mit

$$\sup_{u \in \mathcal{U}} |D_u F_{U|A, \mathbb{I}\{\|\Delta Z\|=0\}\Delta Z, Z_1}(u|a, 0, z)| \leq K_{a,z} < \infty. \quad (73)$$

Hier werden die Differenzierbarkeitsannahmen aus Annahme 3 modifiziert für den Fall der binären Wahl. Obwohl die Indikatorfunktion offensichtlich nicht differenzierbar ist, wird Differenzierbarkeit der bedingten Verteilungsfunktion von U_t verlangt. Die Regularitätsbedingungen sichern ab, dass die Erwartungswerte existieren und dass die entsprechenden Auf- und Ableitungen möglich sind.

Annahme 11:

Es sei $\mathcal{S} := \{s \in \mathbb{R} : s = (x + \xi)' \beta(a), x \in \mathcal{X}, \xi \in \mathcal{N}_\epsilon, a \in \mathcal{A}\}$ und für jedes $(s, a, z) \in \mathcal{S} \times \mathcal{A} \times \mathcal{Z}$ sei

$$\Psi(s, a, z) := 1 - F_{U|A, \mathbb{I}\{\|\Delta Z\|=0\} \Delta Z, Z_1}(-[s + z' \gamma_0 + \alpha(a)]|a, 0, z). \quad (74)$$

Für jedes $(x, z) \in \mathcal{X} \times \mathcal{Z}$ existiert eine μ -integrierbare Funktion $a \rightarrow D(a|x, z)$, so dass

$$\begin{aligned} \sup_{\xi \in \mathcal{N}_\epsilon} |D_{ss} \Psi((x + \xi)' \beta(a), a, z) f(a|\xi, x, 0, z)| &\leq D(a|x, z) \\ \sup_{\xi \in \mathcal{N}_\epsilon} |D_s \Psi((x + \xi)' \beta(a), a, z) D_\xi f(a|\xi, x, 0, z)| &\leq D(a|x, z). \end{aligned} \quad (75)$$

Hier werden die Dominanzbedingungen von Annahme 7 modifiziert um den speziellen Gegebenheiten des binären Modells gerecht zu werden.

Annahme 12:

Die Gewichtungsfunktion $b : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}^+$ ist messbar, so dass $0 < \bar{\psi}_b < \infty$, wobei

$$\bar{\psi}_b := \int \left\{ \int D_s \Psi(x' \beta(a), z, a) F(da|0, x, 0, z) \right\} b(x, z) F(dx, dz) \quad (76)$$

In dieser Annahme wird formal die Gewichtungsfunktion spezifiziert. Dies sichert ab, dass diese ausreichend integrierbar ist.

Annahme 13:

Es sei für jedes $(x, z) \in \mathcal{X} \times \mathcal{Z}$

$$\bar{\psi}(x, z) := \int D_s \Psi(x' \beta(a), z, a) f(a|0, x, 0, z) \mu(da|x, 0, z) < \infty \quad (77)$$

und es werde angenommen, dass

$$\sup_{a \in \mathcal{A}} |D_s \Psi(x' \beta(a), z, a) - \bar{\psi}(x, z)| \leq K_{x,z} < \infty. \quad (78)$$

Annahme 14:

Es sei angenommen, dass

$$\sup_{(a,x,z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}} |D_s \Psi(x' \beta(a), z, a) b(x, z) - \bar{\psi}_b| \leq K_b < \infty. \quad (79)$$

Die letzten beiden Annahmen erlauben es, Grenzen für die marginalen Effekte herzuleiten.

Um das nichtparametrische Identifikationsergebnis für diese Struktur aufzustellen, sei $\beta^*(x, z) := D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0}$ und $\beta_b^* := \mathbb{E}[\beta^*(X_1, Z_1) b(X_1, Z_1)]$. Beide enthalten nur die Verteilung beobachtbarer Zufallsvariablen und sind deshalb empirisch verfügbar.

Theorem 2:¹⁸

Es gelten Annahmen 2, 5, 6 und 9-11.

(i) Dann gilt

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &= \mathbb{E}[D_x \mathbb{P}[Y_1 = 1 | X_1, \Delta X, Z_1, \Delta Z, A] | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z]. \\ &= \int \beta(a) D_s \Psi(x' \beta(a), z, a) f(a | 0, x, 0, z) \mu(da | x, 0, z) \end{aligned} \quad (80)$$

(ii) Es sei angenommen Annahme 12 gelte ebenfalls. Dann gilt

$$\mathbb{E}[\beta^*(X_1, Z_1) b(X_1, Z_1)] = \mathbb{E}(\beta(A) D_s \Psi(X_1' \beta(A), Z_1, A) b(X_1, Z_1)). \quad (81)$$

¹⁸ Für die Herleitung sei auf den Anhang in Hoderlein und White (2009) verwiesen.

(iii) Stattdessen sei angenommen, dass Annahme 13 zusätzlich gelte. Dann gilt

$$|\beta^*(x, z) - \bar{\beta}(x, z)\bar{\psi}(x, z)| \leq K_{x,z} \int |\beta(a) - \bar{\beta}(x, z)| f(a|0, x, 0, z) \mu(da|x, 0, z), \quad (82)$$

wobei

$$\bar{\beta}(x, z) := \int \beta(a) f(a|0, x, 0, z) \mu(da|x, 0, z) \quad (83)$$

(iv) Stattdessen sei weiterhin angenommen, dass Annahme 12 und Annahme 14 zusätzlich gelten. Dann gilt

$$|\beta_b^* - \bar{\beta}\bar{\psi}_b| \leq K_b \mathbb{E}(|\beta(A) - \bar{\beta}|), \quad (84)$$

wobei $\bar{\beta} = \mathbb{E}(\beta(A))$.

Damit ist die Identifikation der durchschnittlichen marginalen Wahrscheinlichkeiten für den allgemeinen Fall einer Grundgesamtheit mit heterogenen Zufallskoeffizienten möglich. In Teil (i) wird gezeigt, dass die Ableitung des bedingten Erwartungswertes die beste Annäherung für die Ableitung der heterogenen Wahrscheinlichkeit für ein Individuum liefert, für welches die Informationen $\sigma(X_1, \Delta X, Z_1, \Delta Z, A)$ vorliegen. Dies hat Ähnlichkeit mit der durchschnittlichen strukturellen Funktion von Blundell und Powell (2004), die das Problem der endogenen binären Wahl über eine Kontrollfunktion lösen. Zu beachten ist, dass hier nur gewichtete Durchschnitte der zu Grunde liegenden Koeffizienten $\beta(A)$ identifiziert werden, wobei auch partiell unbekannte positive Gewichte $D_s \Psi(X_1' \beta(A), Z_1, A) b(X_1, Z_1)$ enthalten sind (siehe Teil (ii)). Da das Interesse normalerweise auf den durchschnittlichen marginalen Wahrscheinlichkeiten liegt, ist dies jedoch nur als kleinere Einschränkung zu betrachten.

Die in (iii) gegebenen Grenzen bilden eine bedeutende Form der Stetigkeit. Insbesondere die letzte Grenze zeigt, dass kleine Mengen von unbeobachteter Koeffizientenheterogenität, wie von $E(|\beta(A) - \bar{\beta}|)$ zu kleinen Abweichungen führen von β_b^* von einer skalierten Version von $\bar{\beta} = \mathbb{E}(\beta(A))$, nämlich $\bar{\beta}\bar{\psi}_b$. Dieses Resultat sichert ab, dass die kleinen Mengen von unbeobachtbarer Heterogenität nicht zu großen Abweichungen der identifizierten β_b^* vom ungewichteten Koeffizientendurchschnitt führen. Dennoch ist der mittlere Zufallskoeffizient nur teilweise identifiziert.

Ist die Grundgesamtheit jedoch homogen in den marginalen Effekten, mit einem additiv korrelierten Fixed Effect, so verschwindet die Einschränkung komplett. Das genaue Ergebnis sieht dann folgendermaßen aus:

Theorem 3:¹⁹

Unter den Annahmen 2, 5, 6 und 9-11 und wenn die Population homogen ist, das heißt für alle $a \in \mathcal{A}$, $\beta(a) = \beta_0$, dann ist β_0 identifiziert als:

$$\beta_0 = D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} / \bar{\psi}(x, z) \quad (85)$$

für jedes (x, z) und infolgedessen auch über die durchschnittliche partielle Ableitung

$$\beta_0 \propto \mathbb{E}[D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1, \Delta Z = 0, Z_1] |_{\xi=0} b(X_1, Z_1)]. \quad (86)$$

Dieses Ergebnis liefert die Identifikation der Koeffizienten β_0 im Paneldatenmodell binärer Wahl. Zu beachten ist, dass die Indexstruktur lediglich eine Identifikation der relativen Effekte zweier Koeffizienten ermöglicht. Prinzipiell sind wieder beliebige Abhängigkeiten zwischen den beobachtbaren Variablen und den unbeobachtbaren Komponenten möglich. Die Annahmen sind sogar schwächer als die standardmäßigen Annahmen zur strikten Exogenität in der Literatur für nichtlineare Modelle. Die Schlüsselschritte für den Beweis des Ergebnisses sind den Schritten im allgemeinen nichtseparablen Fall recht ähnlich. Außerdem ist zu beachten, dass die Erweiterungen, die vorher diskutiert wurden weiterhin möglich sind, insbesondere die Einführung zeitlich verzögerter Regressoren und bedingter Endogenität von X_t .

Erweiterung: Durchschnittseffekte

In Abwesenheit von Zufallskoeffizienten ist ein Natural Sample Counterparts-Schätzer zu

$$\begin{aligned} \beta_0 &\propto \mathbb{E}[D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1, \Delta Z = 0, Z_1] |_{\xi=0} b(X_1, Z_1)] \\ &= \mathbb{E}[\beta^*(X_1, Z_1) b(X_1, Z_1)] \end{aligned} \quad (87)$$

gegeben durch

$$\hat{\beta}_{0,n} = n^{-1} \sum_{i=1}^n \hat{\delta}_n(X_i, Z_i) b(X_i, Z_i), \quad (88)$$

¹⁹ Für die Herleitung sei auf den Anhang in Hoderlein und White (2009) verwiesen.

wobei $\widehat{\delta}_n$ der gleiche lokale Polynom-basierte Schätzer ist, wie in (65) definiert. Dieses $\widehat{\beta}_{0,n}$ ist ein Partieller Durchschnitts-Schätzer, analog zu $\widehat{\delta}_{0,n}$, außer dass nun eine Gewichtung durch $b(X_i, Z_i)$ explizit auftaucht. Wie oben, sichern die Bedingungen von Newey (1994) ab, dass

$$n^{1/2} h_n^\alpha (\widehat{\beta}_{0,n} - \beta_b^*) \rightarrow^d \mathcal{N}(0, \sum_{0,b}), \quad (89)$$

wobei α eine Konstante ist, die von den Dimensionen von X und Z abhängt und $\sum_{0,b}$ eine Kovarianz-Matrix ist, deren spezifische Form von der Wahl von g_p abhängt. Die Ableitung für \sum_0 im Anhang lässt sich unter offensichtlichen Modifikationen anwenden um einen Schätzer für $\sum_{0,b}$ zu erhalten.

Selbst in Gegenwart von Zufallskoeffizienten liefert dieser Koeffizient nützliche Informationen über das Vorzeichen des gesuchten Effekts.

3. Empirieteil

3.1. Immobilienkrise und Liquiditätsbeschränkung der Haushalte

3.1.1. Wirtschafts- und Finanzkrise von 2007

Als Ursprung für die im Jahr 2007 begonnene Wirtschafts- und Finanzkrise gilt der Markt für Wohnimmobilien in den USA: Banken und sonstige Kreditfinanzierer handelten zu unvorsichtig bei der Vergabe von Hypotheken an Haushalte mit geringer Kreditwürdigkeit. Dies ermöglichte es, dass der Häuserbestand in den USA stärker wachsen konnte als die Bevölkerung.²⁰

Begünstigt wurde die Entwicklung zum einen durch die expansive Zinspolitik der amerikanischen Notenbank, denn durch die historisch niedrigen Zinsen fühlten sich viele Haushalte regelrecht zur Kreditaufnahme genötigt. Zum anderen lag es im ausdrücklichen Interesse der amerikanischen Politik, dass möglichst viele Haushalte Zugang zu Wohneigentum erhielten. Um die „ownership society“ verwirklicht zu sehen, wurden staatsnahe Unternehmen wie Fannie Mae und Freddie Mac unterstützt und gezielt Subventionen vergeben.

Zudem befanden sich sowohl die Banken, als auch die breite Öffentlichkeit im Glauben, dass sich der Boom der US-Wirtschaft und insbesondere die positive Entwicklung auf dem Häusermarkt auf längere Zeit fortsetzen würden.²¹ Viele amerikanische Haushalte wähnten sich allein schon durch ihr Wohneigentum für ihr Alter abgesichert. Die Sparquote nahm ab und das Rechtssystem ermöglichte die zusätzliche Beleihung der Wohnimmobilien, so dass der Konsum der Haushalte schneller anstieg, als es die laufenden Einkommen sonst erlaubt hätten. Diese vom Markt für Wohnimmobilien ausgehende, veränderte Risikoeinstellung verbreitete sich schnell weiter, bis schließlich riskante Häuserkredite im Rahmen der Verbriefung marktfähig gemacht wurden und zwischen den Banken sowohl national als auch international gehandelt wurden.

Die Ergebnisse sind allgemein bekannt: Mitte 2006 platzt die Blase auf dem Immobilienmarkt. Die Preise sinken bis April 2009 im Durchschnitt um ca. 34%. Allein im Bereich der Wohnimmobilien beziffert sich der Verlust auf etwa acht Billionen Dollar.²² Der Preisverfall der Immobilien, die in vielen Haushalten fehlenden Einkommen und die ab Ende 2004 wieder steigenden Zinsen sorgen dafür, dass tausende

²⁰ Für Abschnitt 3.1 vergleiche Gordon (2008), Weber (2009) und vor allem Sinn (2010).

²¹ Und das obwohl der Anstieg der Häuserpreise bereits weit über dem lag, was man bisher in der Vergangenheit beobachten konnte (siehe Case-Shiller-Index in 3.1.2).

²² Bei den gewerblichen Immobilien ist der Verlust noch größer!

von Haushalten in die Überschuldung geraten und ihre Hypotheken nicht mehr zurückzahlen können.²³

Infolgedessen müssen zahlreiche namhafte Banken und Versicherer Insolvenz anmelden oder von Regierungen gerettet werden. Auf den Aktienmärkten kommt es zu Verwerfungen. Durch die globale Vernetzung weitet sich die Krise auf fast alle bedeutenden Volkswirtschaften der Welt aus. Unter Immobilienkrisen leiden in Europa vor allem Spanien, Irland und Großbritannien. Produktionssenkungen, steigende Arbeitslosigkeit und Unternehmenszusammenbrüche sind auf der ganzen Welt zu beobachten.²⁴ Der Welthandel bricht ein, einige Staaten ergreifen protektionistische Maßnahmen und Konjunkturprogramme in Milliardenhöhe werden beschlossen. Die Schuldenproblematik einzelner Staaten verschärft sich...

3.1.2. Bisherige Risikomaße für den Immobilienbereich

Bereits vor der Wirtschafts- und Finanzkrise von 2007 existierten einige Indikatoren zur Messung gesamtwirtschaftlicher Risiken aus dem Immobiliensektor. Diese setzen in der Regel an den Immobilienpreisen an. Mittlerweile populär geworden ist der bereits in den 80er Jahren entwickelte Case-Shiller-Index (siehe Abbildung 1). Dieser stellt die Entwicklung der Hauspreise in zehn bzw. zwanzig Metropolregionen in den USA dar.²⁵ Längst schon vor Beginn der Krise zeigten sich anhand dieses Index Hinweise auf das Vorliegen einer Immobilienblase, da die Preise weit über ihre historischen Werte hinausgewachsen waren, bei gleichzeitig gesamtwirtschaftlich wesentlich geringerem Wachstum.

Grundsätzlich gibt es viele weitere Möglichkeiten, Risiken im Immobilienbereich aufzudecken.²⁶ Allgemein herrscht zwar die Auffassung vor, Immobilienpreise seien vor allem durch spezifische regionale und institutionelle Faktoren bestimmt.²⁷ Die Untersuchung des Einflusses von Faktoren wie der Kreditmenge oder dem Geldvolumen ist jedoch zunehmend im Kommen.²⁸

Das Hauptproblem bei den meisten empirischen Untersuchungen stellt die Vielzahl der teilweise nur schwer fassbaren, zu berücksichtigenden Einflussfaktoren dar. Ökonometrische Modelle zur Bestimmung von Fundamentalpreisen versuchen auf der Nachfrageseite Faktoren wie das Pro-Kopf-Einkommen, den Realzins oder demogra-

²³ Allein im August 2008 waren rund 12 Millionen US-Haushalte überschuldet.

²⁴ Laut einer Studie von Deutsche Bank Research beläuft sich die krisenbedingte Minderung der Weltproduktion auf vier Billionen US-Dollar (Vgl. Gräf und Schneider (2009), S. 6).

²⁵ Die Einzelindizes von 10 bzw. 20 Metropolregionen in den USA werden aggregiert.

²⁶ Vgl. Girouard et al. (2006), S.3ff.

²⁷ Hiermit beschäftigt sich beispielsweise der Bereich „spatial econometrics“.

²⁸ Eine Vorreiterstudie für Deutschland stellt die Studie von Belke (2009) dar.

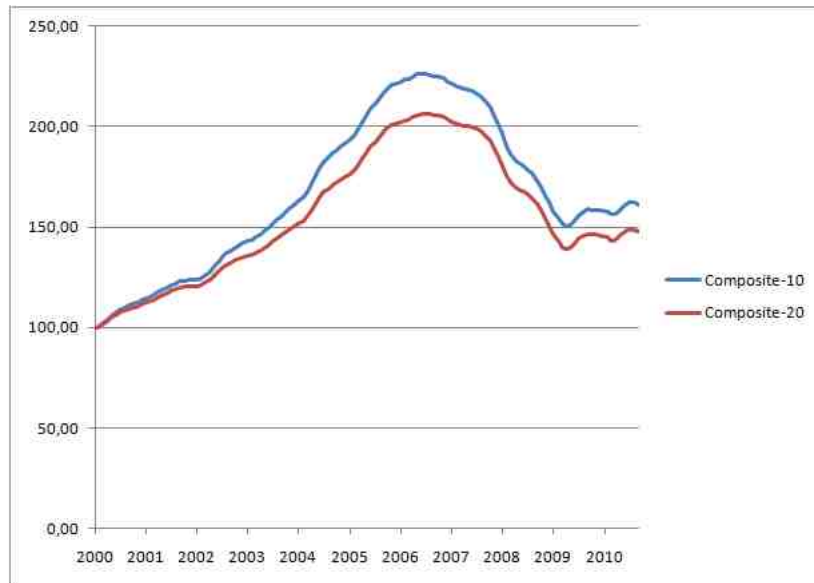


Abbildung 1: Preisentwicklung der US-Immobilien (Case-Shiller-Index, Januar 2000-September 2010).

Quelle: Standard and Poor's, S&P/Case-Shiller Home Price Indices, abgerufen am 9.12.2010.

fische Entwicklungen zu berücksichtigen und auf der Angebotsseite Faktoren wie die Knappheit von Baugrund. Zwischen bestimmten Variablen werden Kointegrationsbeziehungen aufgestellt und die anderen Variablen fließen in einen Fehlerkorrekturterm ein. Die errechneten Schätzungen für die langfristigen Fundamentalpreise können dann mit aktuellen, tatsächlichen Preisen verglichen werden.

Problematisch ist jedoch, dass die Kointegrationsbeziehungen zwischen den Variablen im Zeitverlauf oft nicht stabil sind. Außerdem resultieren je nach Spezifikation ganz unterschiedliche Fundamentalwerte. Die Frage, was ein korrekter Fundamentalwert ist, kann nicht objektiv beantwortet werden und bleibt letztlich immer auch Ermessenssache des Betrachters.

Ein Maß zur Beurteilung der Erschwinglichkeit von Wohneigentum ist die sogenannte Price-to-Income-Ratio.²⁹ Übersteigt diese ihren langfristigen Durchschnitt, so kann dies ein Hinweis auf eine Überbewertung im Immobilienbereich sein. Kurz vor der jüngsten Krise lieferte auch dieser Indikator deutliche Hinweise auf vorhandene Risiken. Dennoch hat auch dieses Maß einige Schwächen, denn langfristig ist die Beziehung zwischen Hauspreisen und Einkommen nicht stabil. Die Kosten für Kredite variieren im Zeitverlauf, außerdem ist das verfügbare Einkommen ein

²⁹ Immobilienpreise im Verhältnis zum durchschnittlich verfügbaren Einkommen.

Durchschnitt über alle verfügbaren Einkommen. Die Teilnehmer auf dem Immobilienmarkt haben jedoch oft höhere Einkommen als der Durchschnitt.

Die Price-to-Rent-Ratio versucht, die Kosten von Wohneigentum ins Verhältnis zu den Mietkosten zu setzen.³⁰ Der Grundgedanke ist, dass ein Gleichgewicht auf dem Häusermarkt genau dann besteht, wenn die erwarteten jährlichen Kosten für ein Haus in Eigentum den Mietkosten für das gleiche Objekt entsprechen. Sind die Kosten für Miete geringer, so ist dies ein Hinweis auf Überbewertung des Wohneigentums. Auch dieses Maß erreichte kurz vor Beginn der Krise in den entsprechenden Ländern historische Höchststände, doch auch hier hängen die Ergebnisse ab von den gewählten Fundamentaldaten, Annahmen über Zinsen und vielen weiteren Faktoren. Oft findet man daher zusätzlich argumentative Ausführungen zur Begründung von Immobilienpreisen. Nur schwer fassbare Faktoren werden dann einfach in die Argumentationsketten aufgenommen. Über die exakten Auswirkungen solcher Faktoren können jedoch lediglich Mutmaßungen getroffen werden.

Was Deutschland insgesamt angeht, so kann festgehalten werden, dass die angesprochenen Indikatoren im Gegensatz beispielsweise zu den USA, bisher keine Hinweise auf Immobilienpreisblasen lieferten. Im Durchschnitt ist in Deutschland etwa seit den 70er Jahren eine Stagnation der realen Immobilienpreise zu verzeichnen.³¹ Zurückführen lässt sich diese Entwicklung auf den in Deutschland vergleichsweise nur geringen Zuwachs der realen verfügbaren Pro-Kopf-Einkommen, das geringe Bevölkerungswachstum und den relativ geringen Urbanisierungsgrad. Hinzu kommen die bereits angesprochenen institutionellen Rahmenbedingungen auf dem Immobilienmarkt.³²

3.1.3. Rolle der Haushalte

Anstatt bei den Immobilienpreisen, besteht jedoch auch die Möglichkeit, direkt bei den Haushalten anzusetzen. Deren wichtige Rolle wurde bereits weit vor den jüngsten Ereignissen erkannt. Insbesondere seit der Deregulierung der Finanzmärkte in den 80er Jahren und dem allgemeinen Rückgang der Zinsen in Folge der niedrigen Inflationsraten, stieg in vielen entwickelten Ländern die private Haushaltsverschuldung an.³³ Daher stellte sich die Frage ob diese Entwicklung eine höhere Sensitivität

³⁰ Berücksichtigt werden sollten dabei sämtliche Risiken, Steuervorteile, Grundsteuern, Instandhaltungen etc.

³¹ Auch wenn dies in einigen Großstädten anders ist.

³² Vgl. Kholodilin, Menz und Siliverstovs (2008), S. 214ff.

³³ Die Finanzierung von Wohneigentum macht in der Regel den größten Anteil an der Haushaltsverschuldung aus.

der Haushalte gegenüber Zins-, Einkommens- und Vermögenspreisänderungen nach sich ziehen würde und ob sich hieraus Gefahren für die Finanzmarktstabilität ergeben würden.³⁴

Zudem ist die finanzielle Situation der privaten Haushalte ein wichtiger Faktor für die Dauer einer Wirtschaftskrise. Verfügen die Haushalte über ausreichend finanzielle Mittel und sind wenig verschuldet, so müssen sie in Krisenzeiten ihre Ausgaben weniger stark reduzieren. Außerdem ist das Ausfallrisiko sowohl für Immobilien-, als auch für Verbraucherkredite geringer.

Alles in allem ist daher das Interesse am Thema Haushaltsverschuldung für die verschiedensten Seiten relevant. Die Deutsche Bundesbank führt im Moment eine großangelegte, länderübergreifende Längsschnittstudie mit der Bezeichnung „Private Haushalte und ihre Finanzen“ durch. Diese soll Aufschluss geben zu Themen wie Sparverhalten oder Altersvorsorge und wertvolle Informationen liefern für die Geld- und Stabilitätspolitik. Aus Sicht der Europäischen Geldpolitik ist zudem die unterschiedliche Sensitivität in den einzelnen Euro-Mitgliedsländern gegenüber ökonomischen Schocks von großem Interesse. Für die Wirksamkeit der monetären Transmissionsmechanismen spielen zudem die institutionellen Unterschiede im Wohnungswesen eine wichtige Rolle.³⁵

Vergleicht man die deutschen Haushalte mit ausländischen Haushalten, zum Beispiel aus den USA, Großbritannien oder den Niederlanden, so kommt man zum Schluss, dass die deutschen Haushalte eine vergleichsweise hohe finanzielle Widerstandskraft aufweisen.³⁶ In Punkto Verschuldung im Verhältnis zum Einkommen stehen die deutschen Haushalte besser da, genauso bei der Verschuldung relativ zum Haushaltsvermögen. Des Weiteren ist die Sparquote in Deutschland im Gegensatz beispielsweise zu Großbritannien oder den USA mit stets knapp über zehn Prozent traditionell recht hoch. Trotz allem gibt es jedoch auch in Deutschland hohe Belastungen vor allem bei jüngeren Haushalten im geringverdienenden Bereich. Die Verschuldung dieser Haushalte muss aufmerksam verfolgt werden.

Ferner zeigt sich bei Vergleichen dieser Art der Zusammenhang zwischen den unterschiedlichen Finanzierungssystemen und der Haushaltsverschuldung. Im marktba-
sierten System in den USA wurden Immobilien bisher teilweise bis zu 100 Prozent be-
liehen. Eine private Haftung der Schuldner über die Immobilien hinaus gab es

³⁴ Vgl. Debelle (2004), S. 59ff.

³⁵ Näheres zum Beispiel bei MacLennan, Muellbauer und Stephens (1998).

³⁶ Vgl. Badunenko, Barasinska und Schäfer (2009), S. 127ff.

nicht. Die tendenziell mit variablen Zinsen ausgestatteten Kreditverträge führten zu einer hohen Sensitivität gegenüber makroökonomischen Schocks. Im bankbasierten deutschen System dagegen war die Vollbeleihung von Immobilien bisher äußerst selten. Der Durchgriff der Gläubiger auf das Privatvermögen der Schuldner war und ist möglich. Zudem hatten die meisten Kredite bisher über längere Zeit fixierte Zinssätze bei einem generell hohen Stellenwert der Eigenfinanzierung. Hohe rechtliche Anforderungen sowie Regelungen wie die Spekulationsfrist von 10 Jahren kommen hinzu. Aus institutioneller Sicht ist damit die vergleichsweise geringere Haushaltsverschuldung in Deutschland nicht verwunderlich. Im Gegensatz zu den USA, kann in Deutschland für die letzten Jahre keine starke Zunahme der Wohneigentumsquote festgestellt werden. Von 1998 bis 2006 nahm diese lediglich um 0,7 Prozentpunkte, von 40,9% auf 41,6% zu.³⁷

Insgesamt wurden bereits viele wichtige Aspekte im Zusammenhang mit der jüngsten Wirtschafts- und Finanzkrise geklärt. Sowohl auf nationaler als auch auf globaler Ebene wurden Konsequenzen gezogen.³⁸ Für die Forschung bleiben jedoch nach wie vor viele Fragestellungen offen, deren Untersuchung gerade im Hinblick auf künftige gesamtwirtschaftliche Entwicklungen, von Bedeutung ist. Hierzu gehört zum Beispiel die Frage, warum in manchen Ländern Immobilienkrisen entstanden sind und in anderen Ländern nicht. Es gibt zwar einige eher argumentative Begründungsansätze, datenbasierte Untersuchungen sind jedoch Mangelware. Zudem ist es geradezu verwunderlich, dass es für die oben als sehr wichtig dargestellte Haushaltsebene kaum ökonometrische Untersuchungen zu den Ursachen und Wirkungen der Verschuldung gibt.³⁹

3.1.4. Beitrag dieser Arbeit

Die vorliegende Arbeit versucht, diese Lücke zu füllen und setzt auf der Haushaltsebene am Zusammenhang zwischen laufendem Haushaltseinkommen und dem Erwerb von Wohneigentum an. Da für den Wohneigentumserwerb in der Regel eine Verschuldung notwendig ist, sollte dem laufenden Einkommen der Haushalte eigentlich eine wichtige Rolle zukommen: Sowohl Kreditgeber als auch die Haushalte selbst können anhand dieser Größe einschätzen, inwieweit die Bewältigung des Schuldendienstes realistisch erscheint.

³⁷ Statistisches Jahrbuch (2010), S. 295.

³⁸ Siehe Weber (2009).

³⁹ Brown und Taylor (2008), S. 642.

Mit Hilfe der im Theorieteil ausführlich dargestellten ökonometrischen Methoden kann der durchschnittliche marginale Effekt des laufenden Einkommens auf die Wahrscheinlichkeit dafür, dass ein Haushalt Wohneigentum besitzt, bestimmt werden.⁴⁰ Entscheidend ist dabei, dass der ökonometrische Ansatz es ermöglicht, für sämtliche zeitkonstante Einflussfaktoren zu kontrollieren. Es spielt keine Rolle, ob diese beobachtbar sind oder nicht. Der Einfluss des laufenden Einkommens lässt sich isolieren. Dies ist von großer Bedeutung, da die Anschaffung von Wohneigentum natürlich nicht nur vom laufenden Einkommen abhängt, sondern auch von einer Vielzahl weiterer Faktoren. Eine direkte Regression der laufenden Haushaltseinkommen auf eine Wohneigentumsvariable wäre erwartungsgemäß stark verzerrt.

Zusätzliche Einflussfaktoren für den Erwerb von Wohneigentum stellen beispielsweise das Alter der Personen im Haushalt, die Kredithistorie des Haushaltes oder das Vermögen des Haushalts dar. Da das Alter in den für diese Arbeit vorliegenden Daten erfasst ist, kann es zusätzlich zum Haushaltseinkommen berücksichtigt werden. Dies ist vor allem deshalb von Bedeutung, weil davon auszugehen ist, dass jüngere Haushalte mobiler sind und sich erst im Laufe der Zeit auf den Erwerb von Wohneigentum einlassen wollen. Bei älteren Haushalten dagegen ist zu erwarten, dass das Interesse am Erwerb von Wohneigentum altersbedingt irgendwann nachlässt, bis es dann irgendwann nicht mehr vorhanden ist.

Die Kredithistorie eines Haushalts ist ein wichtiges Entscheidungskriterium für Kreditgeber. Sie zeigt, inwieweit ein Haushalt in der Vergangenheit fähig war, seine Schulden zu bewältigen und wird zur Prognose künftigen Verhaltens herangezogen. Im Rahmen dieser Analyse ist die Kredithistorie eine unbeobachtbare Variable. Dies stellt jedoch keine Einschränkung dar, denn es handelt sich um einen ziemlich zeitkonstanten Faktor, so wie bereits erläutert. Für die Kredithistorie kann damit kontrolliert werden.

Ebenfalls eine Variable dieser Art ist das ursprüngliche Vermögen eines Haushalts. Dieses kann als konstant angenommen werden und ist somit berücksichtigbar. Wäre dies nicht der Fall, so müsste man mit einer starken Verzerrung der Regressionsergebnisse rechnen, denn beim (ursprünglichen) Haushaltsvermögen handelt es sich um eine Größe, die in der Regel mit dem laufenden Haushaltseinkommen hoch korreliert ist.

⁴⁰ Im Folgenden oft auch abkürzend lediglich als marginale Wahrscheinlichkeit oder als marginaler Effekt bezeichnet.

Inwieweit spielt nun jedoch die marginale Wahrscheinlichkeit, Wohneigentum zu besitzen eine Rolle im Zusammenhang mit der US-Immobilienkrise?

Da angenommen werden kann, dass Haushalte mit geringem bis mittlerem Einkommen normalerweise eher schwer Zugang zu Krediten und damit zu Wohneigentum erhalten, sollte die marginale Wahrscheinlichkeit, Wohneigentum zu besitzen gerade für diese Haushalte positiv sein: Zusätzliches laufendes Einkommen sollte mit einer steigenden Wahrscheinlichkeit, in den Besitz von Wohneigentum zu gelangen, verbunden sein. Bei Hocheinkommenshaushalten dagegen ist zu erwarten, dass diese in Bezug auf ihr laufendes Einkommen wenig liquiditätsbeschränkt sind. Die marginale Wahrscheinlichkeit für den Besitz von Wohneigentum sollte kaum signifikant von null verschieden sein. Zugewinne beim laufenden Einkommen dürften keine Rolle mehr spielen für den Erwerb von Wohneigentum, wenn das Einkommen bereits hoch genug ist.

Der durchschnittliche marginale Effekt des Einkommens auf die Wahrscheinlichkeit dafür, dass ein Haushalt sich im Besitz von Wohneigentum befindet, kann damit als Indikator dafür gesehen werden, wie vorsichtig die Kreditvergabe erfolgt. Änderungen dieses Indikators zeigen an, ob sich die Risikoeinschätzung der Wirtschaftssubjekte bei der Kreditvergabe ändert. Ein Absinken des marginalen Effekts im Zeitverlauf stellt somit ein Indiz dafür dar, dass das laufende Einkommen als Entscheidungskriterium für die Kreditvergabe an Bedeutung verliert. Dies kommt einer sich verringernden, wenn nicht gar verschwindenden, Liquiditätsbeschränkung der Haushalte gleich.

Der Nachweis einer solchen verschwindenden Liquiditätsbeschränkung gelingt Hoderlein und White (2009) in ihrer Forschungsarbeit. Da in der vorliegenden Arbeit eine analoge Untersuchung erfolgt, ist die entscheidende Frage, ob sich der Sachverhalt in Deutschland anders darstellt oder nicht. Die im vorherigen Abschnitt erläuterten Fakten und die Tatsache, dass in Deutschland keine größere Immobilienkrise zu verzeichnen war, sprechen zwar dafür, dass für Deutschland keine verschwindende Liquiditätsbeschränkung gefunden werden kann.⁴¹ Die Frage ob jedoch vielleicht nur eine Verringerung der Liquiditätsbeschränkung deutscher Haushalte nachweisbar ist, bleibt offen und ist gerade im Hinblick auf künftige Risiken im Bereich der privaten deutschen Wohnimmobilien von größtem Interesse.

⁴¹ Ein solches Ergebnis würde die Unterschiede zwischen der amerikanischen und der deutschen Kreditvergabep Praxis verdeutlichen und wäre gleichzeitig mit ein Erklärungsansatz dafür, weshalb Deutschland bisher von einer größeren Immobilienkrise verschont geblieben ist.

3.2. Betrachtung von Hoderlein und White (2009) für die USA

3.2.1. Ökonometrische Umsetzung

Bevor in Abschnitt 3.3 im Detail auf die ökonometrische Untersuchung für Deutschland eingegangen wird, muss zunächst die empirische Arbeit von Hoderlein und White (2009) für die US-Daten knapp dargestellt werden.

Diese schätzen den durchschnittlichen marginalen Effekt des Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen für die beiden Zwei-Perioden-Panel 1999/2001 und 2003/2005. Über den Vergleich der Effekte zeigen sie, in wie weit sich die Einstellung gegenüber dem laufenden Einkommen, als Risikomaß für die Kreditvergabe, verändert hat.

Die theoretische Grundlage stellt Theorem 2 dar:

$$\begin{aligned} & D_\xi \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0} \\ &= \mathbb{E}[D_x \mathbb{P}[Y_1 = 1 | X_1, \Delta X, Z_1, \Delta Z, A] | \Delta X = 0, X_1 = x, \Delta Z = 0, Z_1 = z]. \quad (90) \\ &= \int \beta(a) D_s \Psi(x' \beta(a), z, a) f(a | 0, x, 0, z) \mu(da | x, 0, z). \end{aligned}$$

Für den Besitz von Wohneigentum wird die Binärvariable Y verwendet.⁴² X_1 steht für das Haushalts-Bruttoeinkommen in der ersten Periode des Panels, ΔX steht für dessen Änderung von 1999 auf 2000 bzw. von 2003 auf 2005. Z_1 ist die Variable für das Alter des Haushaltsvorstands in der ersten Periode des jeweiligen Panels, ΔZ stellt dessen Änderung von 1999 auf 2000 bzw. von 2003 auf 2005 dar.

Laut Theorem 2 müssen sowohl Haushalts-Bruttoeinkommen, als auch das Alter des Haushaltsvorstands im betrachteten Zeitraum konstant sein ($\Delta X = 0$ und $\Delta Z = 0$). Tatsächlich ist dies jedoch nicht der Fall. Die Haushalts-Bruttoeinkommen ändern sich im Zeitverlauf. Da diese Änderungen jedoch für den Großteil der Bevölkerung nur sehr gering sind, stellt dies kein größeres Problem dar. Das Alter des Haushaltsvorstands ist ebenfalls nicht konstant: Am Ende des jeweiligen Panels ist jeder Haushaltsvorstand um exakt zwei Jahre älter als am Anfang. Dennoch ist es möglich, als Regressor das Alter in der ersten Periode zu verwenden (alternativ wäre auch das

⁴² $Y=1$: Haushalt besitzt Wohneigentum in der Periode.

$Y=0$: Haushalt besitzt kein Wohneigentum in der Periode.

Abgeleitet wird die Variable aus Angaben darüber ob ein Haushalt Grundsteuer zahlt oder nicht.

im betrachteten Zeitraum durchschnittliche Alter denkbar), denn dieses ändert sich lediglich auf eine deterministische Art und Weise, so dass kein Informationsverlust entsteht. Angenommen werden muss jedoch, dass die Altersänderung in Bezug auf den Erwerb von Wohneigentum einen über alle Altersgruppen hinweg gleichen Effekt hat. Dies ist unrealistisch, da davon auszugehen ist, dass beispielsweise bei einem 30-jährigen die zwei Jahre mehr an Alter zu einer Steigerung der Wahrscheinlichkeit, Wohneigentum zu erwerben führt.⁴³ Bei einem über 50-jährigen dagegen wäre ein viel geringerer Effekt zu erwarten.⁴⁴ Da jedoch nur kleine Zeitintervalle betrachtet werden, lässt sich rechtfertigen, dass man diese Schwäche vernachlässigt. Hierfür spricht zusätzlich, dass der Effekt des Alters des Haushaltsvorstands gar nicht von direktem Interesse ist. Das Alter stellt lediglich eine Kontrollvariable dar. Insgesamt lässt sich daher Theorem 2 anwenden und es gilt $\Delta Z = 0$ und $\Delta X = 0$.

Damit kann die Variable für die Änderung des Wohneigentums regressiert werden auf das Alter des Haushaltsvorstands sowie das Haushalts-Bruttoeinkommen. Konkret geschätzt wird unter Einsatz von Kernel-Methoden die folgende Gleichung:

$$D_{\xi} \mathbb{E}[\Delta Y | \Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z] |_{\xi=0}. \quad (91)$$

Zur Verwendung kommt ein lokaler quadratischer Schätzer mit einem Standard-Epanechnikov-Kern. Die Bandbreite wird so gewählt, dass sie etwas kleiner ist als das kreuzvalidierte Optimum. Änderungen der Bandbreite beeinflussen die Ergebnisse laut Hoderlein und White (2009) jedoch nur marginal. Die Punktschätzer verändern sich kaum.

Für individuenspezifische, jedoch zeitinvariante Variablen (A : z.B. Bildung, Rasse etc.) wird durch die Differenzenbildung kontrolliert. Ebenso wird für persistente Einflussfaktoren, wie zum Beispiel eine schlechte Kredithistorie kontrolliert. Es zeigt sich der Vorteil des nichtparametrischen Ansatzes: Da die Effekte der unbeobachtbaren Variablen nicht von Interesse sind, können viele Variablen ausgelassen werden und der Rechenaufwand sinkt, ohne dass sich die Ergebnisse für die eigentlich interessierenden Variablen verschlechtern.

⁴³ Bei einem jüngeren Menschen sollte mit zunehmendem Alter die Wahrscheinlichkeit, Wohneigentum zu erwerben steigen.

⁴⁴ Bei einem älteren Menschen ist davon auszugehen, dass das Interesse an Wohneigentum altersbedingt irgendwann abnimmt.

3.2.2. Resultate

Für 1999/2001 ergibt sich Abbildung 2:

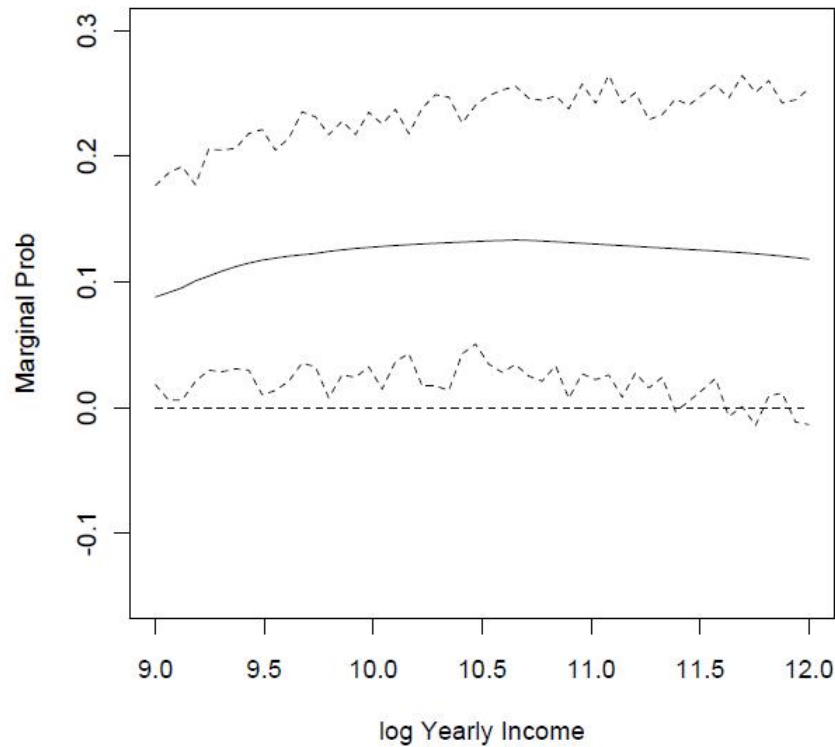


Abbildung 2: Marginale Wahrscheinlichkeit für den Besitz von Wohneigentum bei gegebenem Einkommen (Zeitraum: 1999/2001).

Quelle: Hoderlein und White (2009).

Der marginale Effekt wird dargestellt für ein Alter von $z = 40$. Dieses liegt nahe am PSID-Stichprobendurchschnitt. Mit dem gewählten Darstellungsbereich für die x-Achse sind die Werte der logarithmierten Einkommen für 95% der im Datensatz enthaltenen Individuen erfasst. Die durchgezogene Kurve stellt den Punktschätzer des marginalen Effekts dar. Mit den gestrichelten Linien werden 95%-Bootstrap-Konfidenzbänder ausgewiesen. Klar ersichtlich ist, dass der marginale Effekt im gesamten Bereich (außer evtl. am rechten Rand) positiv und statistisch signifikant von null verschieden ist. Außerdem ist er laut Hoderlein und White (2009) unsensibel in Bezug auf Änderungen der Bandbreite innerhalb plausibler Intervalle. Egal von welchem Haushalts-Bruttoeinkommen man ausgeht, ein höheres Einkommen führt immer zu einer Erhöhung der Wahrscheinlichkeit dafür, dass jemand Wohneigentum besitzt. Alles deutet jedoch darauf hin, dass der positive Effekt bei den unteren bis mittleren Einkommen ausgeprägter ist, denn die Kurve nimmt in diesen Ein-

kommensbereichen einen steigenden Verlauf an. Einkommenszuwächse scheinen sich hier also stärker auf die Wahrscheinlichkeit, Wohneigentum zu erwerben auszuwirken. Bei den höheren Einkommen scheint sich der Effekt abzuschwächen. Die Kurve für die marginalen Effekte nimmt einen fallenden Verlauf an.

Alles in allem entspricht dieses Resultat exakt den theoretischen Vorhersagen, nach denen vor allem die Individuen der unteren und mittleren Einkommensschichten bei der Finanzierung von Wohneigentum einer Liquiditätsbeschränkung unterliegen sollten. Das Resultat liefert somit Evidenz dafür, dass die Vergabe von Hypotheken an Haushalte mit niedrigen bis mittleren Einkommen im Betrachtungszeitraum vorsichtiger erfolgte, als für Haushalte im oberen Einkommensbereich. Zu beachten ist dabei, dass diese Aussage nur im Durchschnitt gelten kann. Im Einzelfall kann sich der Zusammenhang bei einer heterogenen Grundgesamtheit natürlich anders darstellen.

Für den Betrachtungszeitraum 2003/2005 ergibt sich ein etwas anderes Bild (siehe Abbildung 3).

Der Effekt des Einkommens auf die marginale Wahrscheinlichkeit, ein Haus zu besitzen, unterscheidet sich nun nicht mehr signifikant von null. Die durchgezogene Linie befindet sich zwar im positiven Bereich, die beiden 95%-Konfidenzbänder umschließen jedoch über das gesamte Einkommensspektrum hinweg sowohl die null, als auch die durchgezogene Linie. Die einzige mögliche Aussage ist daher, dass sich im betrachteten Zeitraum die durchschnittliche Wahrscheinlichkeit, ein Haus zu besitzen, über das Einkommensspektrum hinweg nicht ändert, wenn man für unbeobachtbare Variablen, wie zum Beispiel eine schlechte Kredithistorie, kontrolliert.

Insgesamt hat laut Hoderlein und White (2009) die Wohneigentumsrate in den USA zwischen 1999 und 2005 von 0,64 auf 0,69 zugenommen. Nimmt man an, dass Haushalte aus den höheren Einkommensschichten ihre Häuser nicht verkauft haben um zur Miete zu wohnen, dann bedeutet dies, dass während des Betrachtungszeitraums insbesondere bei den Haushalten mit niedrigem bis mittlerem Einkommen ein „Catching-up“ in Sachen Wohneigentum stattgefunden hat. Dieses lässt sich damit erklären, dass das laufende Einkommen als Entscheidungsgrundlage für die Kreditvergabe an Bedeutung verloren hat.

In Anhang A3 werden die Unterschiede zwischen den beiden Betrachtungszeiträumen nochmal verdeutlicht. Auf dem 90%-Signifikanzniveau liegt die Funktion für 1999/2001 außerhalb der punktwisen Konfidenzbänder der Funktion für 2003/2005.

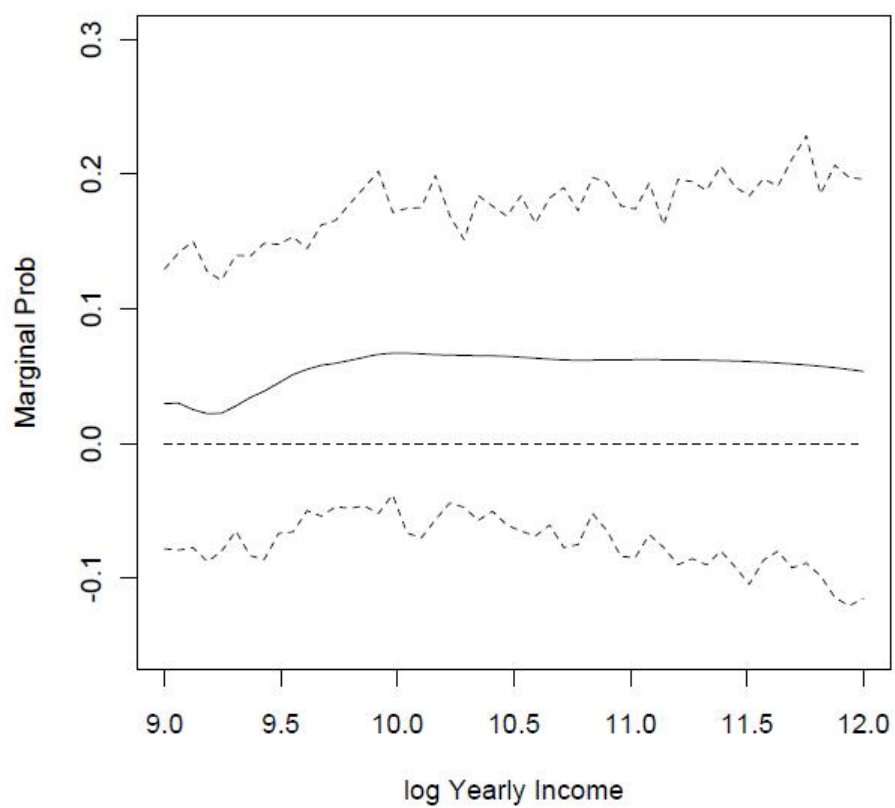


Abbildung 3: Marginale Wahrscheinlichkeit für den Besitz von Wohneigentum bei gegebenem Einkommen (Zeitraum: 2003/2005).

Quelle: Hoderlein und White (2009).

Für das 95%-Niveau ist die Evidenz nicht ganz klar. Zu beachten ist jedoch, dass es sich um punktweise Tests handelt, deren Schwächen allgemein bekannt sind. Vermutet werden kann damit, dass beispielsweise L_2 -Distanztests die Nullhypothese, die Funktionen seien nicht signifikant voneinander verschieden, auf sämtlichen Signifikanzniveaus verwerfen würden, wenn bereits die punktweisen Tests hierzu Anlass liefern.

Insgesamt bestätigen die Ergebnisse die theoretischen Vorüberlegungen voll und ganz. Die sich ändernde Einstellung in den USA, einerseits der Verbraucher bezüglich der Verschuldung, andererseits der Kreditinstitute in Bezug auf die Kreditvergabe, wird von den Daten klar und deutlich widergegeben. Auf dem Weg in eine „ownership society“ spielten Einkommensdifferenzen und Einkommensrisiken keine Rolle mehr für die Kreditvergabe. Die Liquiditätsbeschränkung der Haushalte hat sich gelockert. Betont sei jedoch, dass sich diese Aussage lediglich in Bezug auf das laufende Einkommen treffen lässt. Dieses wurde durch den ökonometrischen Ansatz von sämtlichen anderen zeitkonstanten Faktoren isoliert.

Des Weiteren sei erwähnt, dass Hoderlein und White (2009) mit einigen Datenveränderungen experimentieren. Sie untersuchen beispielsweise die Auswirkungen davon, keine Außereißerdaten mit einzubeziehen oder lediglich eine Stichprobe von Individuen mit konstant guter Gesundheit zu betrachten. Insgesamt ergeben sich jedoch keine bedeutsamen Ergebnisänderungen. Lediglich bei der Betrachtung der Untergruppe der nicht-afro-amerikanischen Bevölkerung scheinen sich geringere Effekte zu ergeben. Aufgrund mangelnder Beobachtungen ist dies jedoch laut Hoderlein und White (2009) sehr vage.

3.3. Vergleichende Untersuchung für Deutschland

3.3.1. Datengrundlage

Wie bereits eingangs erwähnt, stammen die in der Analyse für Deutschland verwendeten Daten aus dem Sozio-ökonomischen Panel. Dieses wird bereits seit 1984 vom deutschen Institut für Wirtschaftsforschung (DIW) in Berlin betrieben und besteht aus jährlichen, repräsentativen Befragungen unter privaten deutschen Haushalten, wobei die befragten Personen über 17 Jahre alt sein müssen. Die erfassten Daten reichen von Persönlichkeitsmerkmalen, über Einkommensdaten, bis hin zu ganzen Familienbiografien.⁴⁵ Die bisher erfasste Datenmenge ist enorm. Allein für das Jahr 2008 liegen 297 verfügbare Datensätze vor. Diese umfassen knapp fünf Millionen Daten-Zeilen mit 39.550 Beobachtungen. Das SOEP stellt damit eine hervorragende Grundlage für die unterschiedlichsten Forschungsarbeiten dar. Die Themenfelder entstammen hauptsächlich den Wirtschafts- und Sozialwissenschaften, denn gesellschaftliche Fragestellungen lassen sich mit Hilfe des SOEP besonders gut untersuchen. Ein großer Teil der durchgeführten Studien ist damit auch höchst politikrelevant und liefert wertvolle Erkenntnisse und Lösungsmöglichkeiten für aktuelle Probleme.⁴⁶ Nicht zu vernachlässigen ist jedoch, dass mittlerweile auch viele andere Fachgebiete, wie zum Beispiel die Energie- oder Umweltwissenschaften, mehr und mehr vom SOEP Gebrauch machen. Die Gesamtzahl der SOEP-Studien lag bis zum Jahr 2008 bei rund 5.000.⁴⁷

Besonders reizvoll für die Forschung ist, dass sich das Potenzial des SOEP mit jedem weiteren Jahr, in dem Befragungen durchgeführt werden, noch mehr entfalten kann. Die bisherigen Forschungsansätze können auf der sich stetig vergrößernden Datengrundlage immer wieder neu angewendet werden. Durch zahlreiche Erweiterungen und Ergänzungen der Stichproben (siehe Tabelle 1) können außerdem zusätzliche Analysen bestimmter gesellschaftlicher Gruppen erfolgen (beispielsweise Hocheinkommenshaushalte⁴⁸). In Zukunft soll die Verfolgung der kompletten Lebensverläufe ganzer Kohorten ermöglicht werden. Ein weiterer interessanter Aspekt ist die Einbettung des SOEP in eine weltweite Forschungsinfrastruktur mit Panel-

⁴⁵ Genaueres zu den Variablen beispielsweise bei Haisken-DeNew und Frick (2005).

⁴⁶ Siehe beispielsweise SOEP Wave Report (Headey und Holst (2008)): Die behandelten Themen reichen von Kinderarmut über Arbeitsmarktreformen und Lebenszufriedenheit bis hin zu Lohnunterschieden zwischen Mann und Frau.

⁴⁷ Vgl. Wagner et al. (2008), S. 301ff.

⁴⁸ Eine in diesem Zusammenhang noch laufende Studie, für die das SOEP relevant ist, wird im Moment vom Institut für angewandte Wirtschaftsforschung in Tübingen durchgeführt. Der Titel lautet: Möglichkeiten und Grenzen der Reichtumsberichterstattung. Auftraggeber ist das Bundesministerium für Arbeit und Soziales.

daten zu Haushalten und Personen. Um internationale Vergleiche zu ermöglichen, wurde das „Cross National Equivalent File“ (CNEF) geschaffen. Insgesamt sind in diesem bis dato rund zwei Millionen Beobachtungseinheiten erfasst, die über die Zeit und über die Länder hinweg verglichen werden können. Die beteiligten Länder sind Großbritannien (British Household Panel Survey), die USA (Panel Study of Income Dynamics), Kanada (Survey of Labour and Income Dynamics), Australien (Household Income and Labour Dynamics Survey), die Schweiz (Schweizer Haushaltspanel) und Deutschland (SOEP).⁴⁹

Tabelle 1: Stichproben im SOEP

Stichprobe A	(Deutsche) Haushalte in der BRD Hauptstichprobe, Start 1984 N=4.528 realisierte Haushalte
Stichprobe B	Ausländische Haushalte in der BRD Start 1984 N=1.393 realisierte Haushalte
Stichprobe C	Privathaushalte in der DDR Start im Juni 1990 N=2.179 realisierte Haushalte
Stichprobe D	Zuwanderer-Privathaushalte in Deutschland Start 1994/95 N=374 realisierte Haushalte
Stichprobe E	Haushalte in Deutschland Ergänzungsstichprobe, Start 1998 N=1.056 realisierte Haushalte
Stichprobe F	Haushalte in Deutschland Ergänzungsstichprobe, Start 2000 N=6.052 realisierte Haushalte
Stichprobe G	Hocheinkommens-Privathaushalte in Deutschland Hocheinkommensstichprobe, Start 2002 N=1.224 realisierte Haushalte
Stichprobe H	Haushalte in Deutschland Ergänzungsstichprobe, Start 2006 N=1.506 realisierte Haushalte

Quelle: Wagner et al. (2008).

⁴⁹ Vgl. Frick et al. (2008), S. 110ff.

3.3.2. Datenaufbereitung

Tabelle 2: Die wesentlichen Variablen

Variable	Mögliche Ausprägungen
Änderung des Wohneigentums innerhalb des betrachteten Zeitraums	0: keine Änderung 1: Erwerb von Wohneigentum -1: Verlust von Wohneigentum (benötigt für jedes betrachtete Panel)
Alter des Haushaltsvorstands im ersten Jahr des betrachteten Zeitraums	beliebige Werte ab 18 Jahren (benötigt für jedes betrachtete Panel)
Jährliches Haushalts-Nettoeinkommen	beliebige Werte ab 0 Euro aufwärts (benötigt für jedes einzelne Jahr)

Quelle: Eigene Darstellung.

Für die ökonometrischen Schätzungen werden zwei Datensätze gebildet: Ein Zwei-Perioden-Panel für die Jahre 1999/2001 und eins für die Jahre 2003/2005.⁵⁰ Die Individuen entstammen damit den Stichproben A bis E. Die gewählten Jahre entsprechen genau denen, die auch Hoderlein und White (2009) wählen. Vergleiche mit deren Ergebnissen werden so erleichtert.

Zunächst werden aus den ursprünglichen SOEP-Daten die Daten für die Haushalts-Nettoeinkommen der Jahre 1999, 2001, 2003 und 2005 entnommen. Bei Hoderlein und White (2009) werden zwar die Haushalts-Bruttoeinkommen verwendet, jedoch sind diese im SOEP nicht direkt verfügbar. Es bleibt daher nichts anderes übrig, als die Netto-Werte zu verwenden. Da das Ziel dieser Analyse eher auf dem Nachweis grundsätzlicher Effekte liegt und weniger auf der Berechnung marginaler Effekte auf ganz bestimmten Einkommensniveaus, ist dies jedoch nicht als große Einschränkung zu betrachten. Für exakte Vergleiche mit den Ergebnissen von Hoderlein und White (2009) müsste dies jedoch berücksichtigt werden.

Des Weiteren ist problematisch, dass im SOEP für jedes Jahr nur ein monatliches Nettoeinkommen pro Haushalt ausgewiesen wird. Da jedoch das jährliche Nettoeinkommen benötigt wird, müssen die Monatswerte mit zwölf multipliziert werden. Vernachlässigt wird dabei jedoch, dass die Befragungen, in denen die Einkommen erfasst werden, nicht nur an einem Tag im Jahr, sondern über mehrere Monate hinweg

⁵⁰ Die benötigten Variablen sind in Tabelle 2 dargestellt.

erfolgen. Damit wäre es theoretisch möglich, dass die mit zwölf multiplizierten monatlichen Einkommen, die am Anfang des Jahres erfasst werden, das durchschnittliche jährliche monatliche Einkommen unterschätzen, während Einkommenswerte, die am Ende des Jahres erfasst werden dieses überschätzen. Da man jedoch von einer Zufallsstichprobe ausgeht und lediglich durchschnittliche Effekte von Interesse sind, kann dieses Problem ignoriert werden: Man geht davon aus, dass die Einkommenswerte im Mittel korrekt erfasst sind.

Für die späteren Regressionen werden die jährlichen Haushalts-Nettoeinkommen außerdem logarithmiert. Die Verteilung der Werte ist dann der Normalverteilung näher (siehe Abbildung 4). Dies ist für die späteren Analysen günstiger. Was die Verteilung der Einkommen betrifft, so zeigt sich wenig überraschend, dass im Jahr 1999 der absolute Großteil der logarithmierten Haushalts-Nettoeinkommen zwischen neun und elf lag. Dies entspricht dem Bereich von 10.000 bis 60.000 Euro.⁵¹

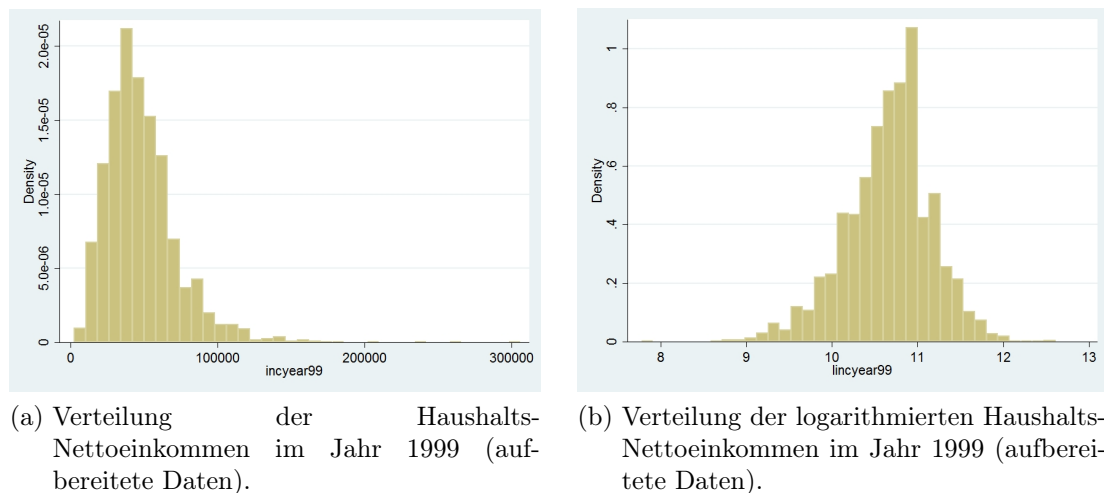


Abbildung 4: Verteilung der Haushalts-Nettoeinkommen 1999.

Quelle: Eigene Darstellung.

Für die Altersvariable werden die im SOEP vorhandenen Daten zu den Geburtsjahren der jeweiligen Haushaltsvorstände verwendet. Anhand dieser Daten können die jeweiligen Alterswerte berechnet werden. Als Bezugsjahre werden 1999 und 2003 gewählt. Dies sind die jeweils ersten Jahre der beiden Zwei-Perioden-Panel. Denkbar wäre jedoch genauso das mittlere oder letzte Jahr des jeweiligen Panels. Da Hoderlein und White (2009) auch das erste Jahr wählen, wird hier deren Methodik gefolgt.

⁵¹ Für die restlichen Jahre ist dies ganz ähnlich (siehe Anhang A1).

Bei der Verteilung der Altersvariable fällt auf, dass zwar der Mittelwert 1999 in etwa bei 48 Jahren liegt.⁵² Die meisten Werte liegen jedoch innerhalb zweier Bereiche, die grob etwa zwischen 25 und 40 Jahren sowie 50 und 70 Jahre liegen (siehe Abbildung 5).

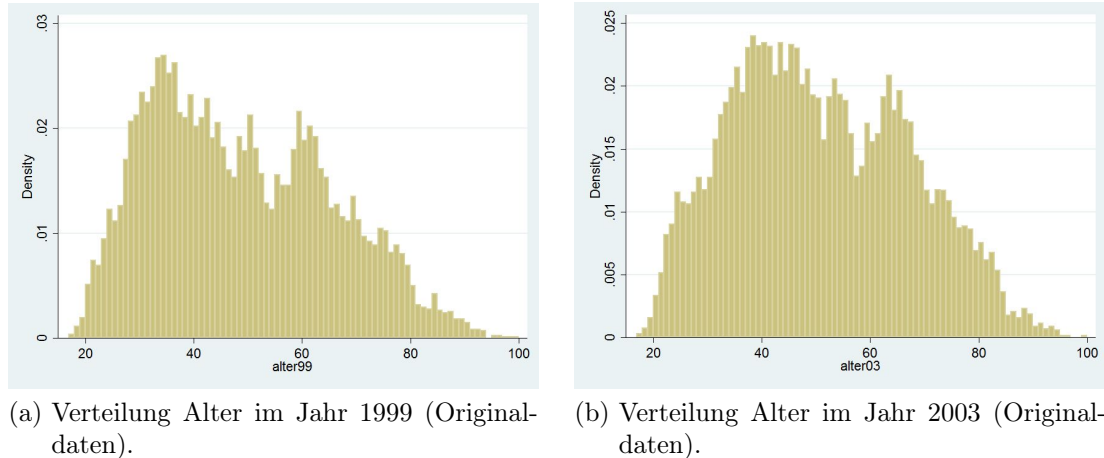


Abbildung 5: Verteilung der Variable Alter.
Quelle: Eigene Darstellung.

Was das Wohneigentum angeht, so gibt es im SOEP für jedes Jahr eine Variable, die Angaben über die Wohnsituation des Haushalts liefert: Zwischenmiete, Hauptmiete oder Eigentum. Hieraus kann die benötigte Binärvariable generiert werden, die anzeigt ob Eigentum vorliegt oder nicht. Um nun die Änderung des Eigentumsstatus während des Betrachtungszeitraums darzustellen, wird die Differenz für die Eigentumsvariable zwischen der letzten und der ersten Periode des jeweiligen Panels berechnet. Die resultierende Variable gibt dann an, ob der jeweilige Haushalt im Betrachtungszeitraum Eigentum erworben oder verloren hat, bzw. ob sich nichts geändert hat (Siehe Tabelle 3 und 4).

Die Zusammenführung der jeweiligen Daten erfolgt anhand eines gemeinsamen Identifikators. Verwendet wird die für das Jahr 1999 gegebene aktuelle Haushaltsnummer⁵³. Durch Elimination aller Beobachtungen, für die in einer der Perioden Werte fehlen, wird sichergestellt, dass eine Stichprobe vorliegt, für die die benötigten Variablen auch in allen vier Perioden vorhanden sind. Insgesamt verbleiben 4.787 Haushalte.

⁵² Das Durchschnittsalter liegt somit im SOEP um etwa acht Jahre höher als im PSID.

⁵³ Nicht zu verwechseln mit der Ursprungshaushaltsnummer. Diese entspricht nur im ersten Jahr der aktuellen Haushaltsnummer.

Tabelle 3: Änderung des Eigentumsstatus zwischen 1999 und 2001

Änderung Eigentum 99/01	Abs. Häufigkeit	Rel. Häufigkeit (%)
-1	59	1,23
0	4.549	95,03
1	179	3,74
Total	4.787	100,00

Quelle: Eigene Darstellung.

Tabelle 4: Änderung des Eigentumsstatus zwischen 2003 und 2005

Änderung Eigentum 03/05	Abs. Häufigkeit	Rel. Häufigkeit (%)
-1	59	1,23
0	4.566	95,38
1	162	3,38
Total	4.787	100,00

Quelle: Eigene Darstellung.

Was die Repräsentativität der ursprünglichen Daten angeht, so kann davon ausgegangen werden, dass diese weitgehend gewährleistet ist. Da jedoch keine Hochrechnung der Daten auf die Grundgesamtheit aller Personen in Deutschland erfolgt, sind kleinere Verzerrungen nicht auszuschließen.⁵⁴ Da es in dieser Arbeit jedoch mehr um durchschnittliche Effekte für die Gesamtbevölkerung und weniger um die Effekte für einzelne Gruppen geht, kommt diesem Aspekt nur eine untergeordnete Rolle zu.

Bei der Aufbereitung der Daten fallen allerdings zahlreiche Beobachtungen heraus, denn nicht für alle Individuen liegen Werte über alle vier Perioden vor. Teilweise gibt es auch unplausible Fehlangaben, für die bereinigt werden muss. Aus dem Vergleich der deskriptiven Statistiken für die ursprünglichen Daten aus den jeweiligen Perioden des SOEP mit den deskriptiven Statistiken der aufbereiteten Daten, kann jedoch geschlossen werden, dass sich durch die Bearbeitung der Daten keine tiefgreifenden Veränderungen in den Verteilungen der Variablen ergeben.⁵⁵ Lediglich die Rolle von Ausreißern scheint sich etwas zu verringern, da einige dieser Werte wegfallen (siehe beispielsweise Einkommen in 2003 oder 2005). Ansonsten sind sich die Verteilungen

⁵⁴ Theoretisch können mit Hilfe spezieller Verfahren unterschiedliche Auswahlwahrscheinlichkeiten für verschiedene Merkmale in den jährlichen Stichproben berücksichtigt werden.

⁵⁵ In Anhang A1 sind die deskriptiven Statistiken für die ursprünglichen Daten dargestellt. Die Statistiken für die aufbereiteten Daten können Anhang A2 entnommen werden.

sehr ähnlich. Dies ist ein wichtiges Ergebnis, denn somit wurden bei der Aufbereitung der Daten keine unbeabsichtigten selektiven Änderungen vorgenommen. Es kann also davon ausgegangen werden, dass die aufbereiteten Daten weitgehend genauso repräsentativ sind wie die ursprünglich gegebenen Daten.

3.3.3. Schätzung

Die Untersuchung der deutschen Daten erfolgt, wie bereits erwähnt, ganz analog zur Untersuchung bei Hoderlein und White (2009).⁵⁶ Dennoch ergeben sich einige Unterschiede, die erläutert werden müssen. Mit Hilfe von Kernel-Methoden wird für die deutschen Daten ebenfalls folgende Gleichung geschätzt:

$$D_{\xi}\mathbb{E}[\Delta Y|\Delta X = \xi, X_1 = x, \Delta Z = 0, Z_1 = z]|_{\xi=0}. \quad (92)$$

Regressiert wird nun die Binärvariable für die Änderung des Wohneigentums (ΔY) auf das Haushalts-Nettoeinkommen in der ersten Periode des Panels (X_1) sowie auf das Alter des Haushaltsvorstands, ebenfalls in der ersten Periode (Z_1). Für individuenspezifische, konstante und persistente Variablen A wird wieder über die Durchschnittsbildung kontrolliert. Wie bei Hoderlein und White (2009) wird ein lokaler quadratischer Schätzer mit einem Standard Epanechnikov-Kern gewählt. Für die Bandbreite werden allerdings nun die jeweils optimalen Werte gewählt, da sich Experimente mit verschiedenen Bandbreiten als wenig nützlich gezeigt haben.

Da streng genommen wieder $\Delta X = 0$ und $\Delta Z = 0$ gelten muss, fragt sich, wie die in 3.2.1 erläuterte Problematik bei den deutschen Daten gehandhabt wird.

Was das Alter des Haushaltsvorstands angeht, so behilft man sich mit der gleichen Annahme wie bei Hoderlein und White (2009) und verwendet einfach das Alter in der ersten Periode des Panels. Per Definition legt man fest, dass gilt $\Delta Z = 0$.

Um zu untersuchen, inwieweit sich die Ergebnisse für verschiedene Altersgruppen unterscheiden, werden verschiedene Altersgruppen für die Regressionen gebildet (siehe Tabelle 5).⁵⁷

Bei den Haushalts-Nettoeinkommen bietet es sich an, nicht nur diejenigen Haushalte zu betrachten, deren Einkommen innerhalb der jeweiligen Panel-Zeiträume konstant bleiben,⁵⁸ sondern auch Haushalte mit Einkommensänderungen innerhalb gewisser Bandbreiten (siehe Tabelle 6). Zwar kann hierdurch Theorem 2 nicht mehr exakt

⁵⁶ Vgl. Abschnitt 3.2.1.

⁵⁷ Näheres dazu in 3.3.4.

⁵⁸ Für das Panel 1999/2001 sind dies 401 Haushalte, für das Panel 2003/2005 sind es 457 Haushalte.

erfüllt werden, jedoch kann dies in Kauf genommen werden, da sich zum einen die Zahl der Beobachtungen erhöht und zum anderen die späteren Untersuchungen zeigen, dass sich die Ergebnisse selbst dann kaum verändern, wenn Haushalte mit starken Einkommensänderungen mit einbezogen werden.⁵⁹

Tabelle 5: Untersuchte Altersgruppen

Altersgruppe	„jung“	„alt“	„um den Durchschnitt“	„alle“
Alter	25-40	50-70	41-55	0-...

Quelle: Eigene Darstellung.

Tabelle 6: Untersuchte Einkommensänderungen

Änderung Haushalts-Nettoeinkommen	1.	2.	3.	4.
Bandbreite in Euro	+/- 0	+/- 750	+/- 3.000	<i>alle</i>

Quelle: Eigene Darstellung.

3.3.4. Ergebnisse

Eine ausführliche Darstellung der Ergebnisse in grafischer Form, findet sich für das Panel 1999/2001 in Anhang A.5. und für das Panel 2003/2005 in Anhang A.6.

Da eine Vielzahl von Grafiken ausgewiesen wird, fasst Anhang A.4. mit zwei Tabellen die wesentlichen Informationen noch einmal zusammen.

Da die ausführliche Erläuterung aller einzelnen Grafiken wenig zielführend ist, werden im Nachfolgenden einzelne Grafiken exemplarisch herausgegriffen. Anhand dieser werden die Ergebnisunterschiede verschiedener Varianten der ökonometrischen Schätzung erläutert und es wird versucht, inhaltliche Schlussfolgerungen zu ziehen.

Was die Grafiken für die deutschen Schätzergebnisse angeht, so ist zu beachten, dass der Darstellungsbereich (bzw. das angezeigte Einkommensspektrum) nicht für alle Grafiken gleich ist. Bei direkten Vergleichen zwischen einzelnen Grafiken ist Vorsicht geboten. Der Bereich, innerhalb dessen die meisten Haushalts-Nettoeinkommen liegen, liegt zwischen 9 und 11 (logarithmierte Einkommenswerte). Dies entspricht in

⁵⁹ Bei nichtparametrischen Verfahren ist eine große Zahl von Beobachtungen besonders wichtig.

etwa 10.000 bis 60.000 Euro. Die durchgezogene Kurve stellt wie bei Hoderlein und White (2009) den Punktschätzer für die durchschnittliche marginale Wahrscheinlichkeit für den Besitz von Wohneigentum dar. Die grau eingefärbte Fläche stellt den von den 95%-Konfidenzbändern umschlossenen Bereich dar. Deutlich erkennbar ist die Weitung der Konfidenzbänder an den äußeren Rändern der Darstellungsbe-
reiche. Diese ist damit zu erklären, dass es nur wenige Haushalte mit Einkommen in diesen Bereichen gibt. An den Rändern liegen weniger Beobachtungen. Dies geht zu Lasten der Präzision der Schätzung innerhalb dieser Bereiche. Wirklich Aussagekräftig ist die Kurve für die durchschnittliche marginale Wahrscheinlichkeit daher lediglich in Bereichen, in denen die Konfidenzbänder relativ nahe an den Kurven liegen. Ist im Folgenden von einem signifikant positiven Effekt bzw. einer signifikant positiven marginalen Wahrscheinlichkeit die Rede, so ist damit gemeint, dass die Kurve der marginalen Wahrscheinlichkeit samt ihren Konfidenzbändern auf dem relevanten Einkommensbereich an mindestens einem Punkt oberhalb von null liegt.

Bezüglich der Kontrollvariable „Alter des Haushaltsvorstands“ ist anzumerken, dass die Schätzung mit den deutschen Daten es leider nicht erlaubt, die marginale Wahrscheinlichkeit für lediglich ein bestimmtes Alter auszuweisen.⁶⁰ Verwendet man nur die Daten von Haushalten, die das Durchschnittsalter der Haushaltsvorstände in der Stichprobe aufweisen, so bleiben nur noch wenige Beobachtungen für die Regression übrig. Selbst bei Wahl des häufigsten Alters bleibt dieses Problem bestehen. Mit der gewählten Regressionsmethode kann leider nur eine erklärende Variable verwendet werden. Hinzu kommt, dass für die ökonometrischen Schätzungen noch die Bandbreiten bezüglich der Einkommensänderungen berücksichtigt werden müssen. Damit sinkt die Zahl der Beobachtungen noch weiter.

Als Ausweg bietet sich die Verwendung der bereits im vorangegangenen Abschnitt eingeführten Altersgruppen an. Dies hat den positiven Nebeneffekt, dass Vergleiche über die Altersgruppen hinweg ermöglicht werden. Der Nachteil liegt jedoch auf der Hand: Möchte man den durchschnittlichen marginalen Effekt schätzen, so muss man implizit annehmen, dass dieser für alle Individuen einer Altersgruppe gleichermaßen gilt. Dies stellt eine starke Vereinfachung dar, denn für einen Haushalt mit einem 30-jährigen Vorstand stellt sich der durchschnittliche marginale Effekt mit Sicherheit anders dar, als für einen Haushalt mit einem 40-jährigen Vorstand. Da das Alter des Haushaltsvorstands jedoch weiterhin lediglich als Kontrollvariable dient und nur die grundsätzlichen Effekte von Interesse sind, kann diese Vereinfachung gerechtfertigt werden.

⁶⁰ Hoderlein und White (2009) wählen $z=40$.

tigt werden. Die im Folgenden ausgewiesenen durchschnittlichen marginalen Effekte beziehen sich somit nur auf die einzelnen Altersgruppen und sollten nicht auf einzelne Altersstufen bezogen werden. Was die Auswahl der Altersgruppen angeht, so wurde hierzu die Verteilung der Altersvariable untersucht. Die Intervalle wurden so gewählt, dass innerhalb der Gruppen möglichst viele Beobachtungen vorhanden sind.⁶¹

Alles in allem liegt bei der Untersuchung der deutschen Daten stets der Trade-Off zwischen dem genauen Erfüllen der theoretischen Voraussetzungen aus Theorem 2 und der Verwendung einer möglichst großen Beobachtungszahl vor. Wie die späteren Untersuchungen jedoch zeigen, ist dies meist weniger problematisch, als zu Beginn angenommen. Die Schätzungen unter Verwendung unterschiedlich vieler Beobachtungen führt meist zu ganz ähnlichen Ergebnissen.

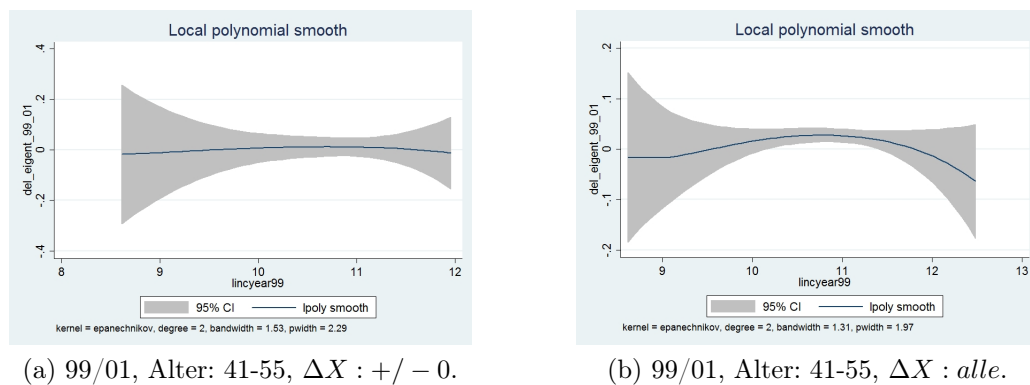


Abbildung 6: Ergebnisse für die Gruppe „um Durchschnitt“ 1999/2001.
Quelle: Eigene Darstellung.

Am ehesten vergleichbar mit den Ergebnissen von Hoderlein und White (2009), sollten die Ergebnisse für die Gruppe „um den Durchschnitt“ (von 41 bis 55 Jahren) sein. Betrachtet man die Ergebnisse für 1999/2001 in Abbildung 6, so zeigt sich, dass unter Berücksichtigung der Bedingung, dass das Haushalts-Bruttoeinkommen konstant bleibt ($\Delta X : +/ - 0$), kein signifikant von null verschiedener marginaler Effekt des Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, gefunden werden kann (siehe Abbildung 6a). Das Konfidenzband umschließt auf dem gesamten dargestellten Einkommensspektrum die x-Achse. Zu beachten ist jedoch, dass für diese Regression lediglich knapp über hundert Beobachtungen verwendet werden können. Das Ergebnis kann damit zwar in die Analyse mit einbezogen werden, sollte

⁶¹ Siehe Tabelle 5 sowie Anhang A2.

jedoch als kritisch betrachtet werden. Lässt man beliebige Einkommensänderungen zu, so lässt sich eine positive marginale Wahrscheinlichkeit finden (siehe Abbildung 6b). Die Zahl der genutzten Beobachtungen liegt hier mit 1.328 wesentlich höher. Streng genommen verstößt man jedoch wieder gegen die theoretischen Vorgaben aus Theorem 2.

Insgesamt ist das Hauptproblem bei der Gruppe „um den Durchschnitt“, dass sich die meisten Beobachtungen leider nicht direkt um den Durchschnitt herum befinden. Die Altersgruppe ist damit eher weniger geeignet für die Analyse, da entweder ein Verstoß gegen die theoretischen Vorgaben oder eine sehr geringe Zahl von Beobachtungen in Kauf genommen werden müssen.

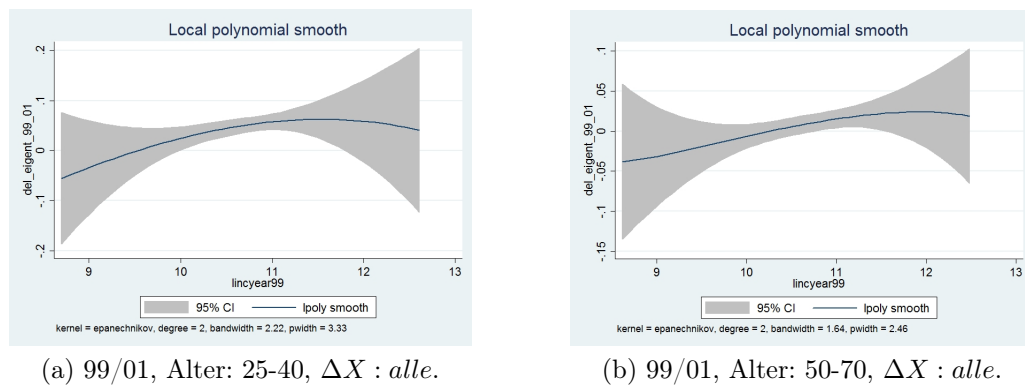


Abbildung 7: Ergebnisse für die Gruppen „jung“ und „alt“ 1999/2001.
Quelle: Eigene Darstellung.

Bei den Gruppen „alt“ und „jung“ sind die im Vorangegangenen erläuterten Probleme ebenfalls vorhanden. Die Situation stellt sich jedoch insgesamt als für die Analyse günstiger dar. Interessanterweise zeigt sich, dass für die Altersgruppe „jung“ die geschätzte Funktion inklusive ihrer Konfidenzbänder in jeder Schätzvariante zumindest teilweise oberhalb von null liegt. Ein signifikanter Effekt des laufenden Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, kann also gefunden werden, egal welche Einschränkung man bezüglich der Einkommensänderungen wählt. Bei der Gruppe „alt“ dagegen ist eine leicht positive marginale Wahrscheinlichkeit nur dann feststellbar, wenn man sämtliche Einkommensänderungen zulässt (siehe Abbildung 7).

Dieses Ergebnis spricht für die Hypothese, dass das laufende Einkommen für jüngere Haushalte einen größeren Effekt auf die Wahrscheinlichkeit, Wohneigentum zu besit-

zen hat. Bei älteren Haushalten ist der Effekt geringer, da das Interesse am Erwerb von Wohneigentum altersbedingt abnimmt. Einkommenszuwächse haben geringere Auswirkungen. Die unterschiedliche Wirkung von Einkommenszuwächsen bei jungen und älteren Haushalten kann damit als erstes wichtiges Ergebnis festgehalten werden.

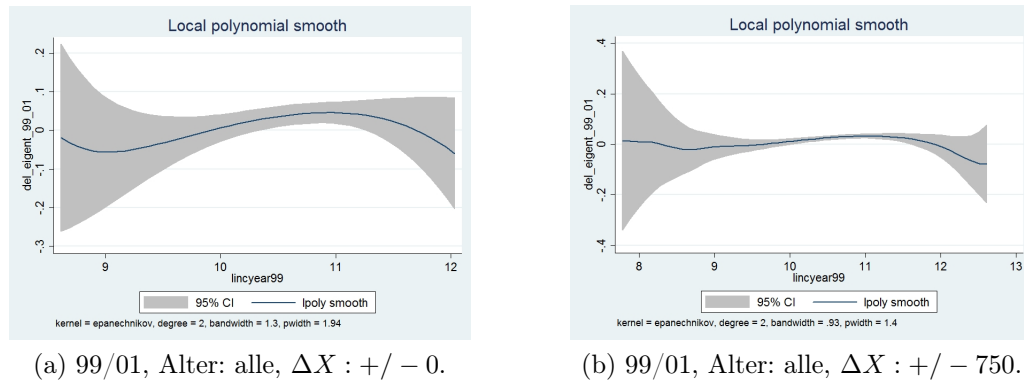


Abbildung 8: Ergebnisse für die Gruppe „alle“ 1999/2001.
Quelle: Eigene Darstellung.

Vernachlässigt man die Kontrolle für das Alter ganz und verwendet die Daten zu Haushalten aller Altersgruppen, so muss man zwar den stärksten Verstoß gegen die theoretischen Vorgaben in Kauf nehmen, jedoch hat man dann die größte Zahl an Beobachtungen zur Verfügung. Ohne Einschränkung bezüglich der Einkommensänderungen kann man auf die Beobachtungen sämtlicher Haushalte zurückgreifen (4.787). Es ergibt sich ein durchweg positiver durchschnittlicher marginaler Effekt des Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen (Siehe Abbildung 8).

Insgesamt kann damit für die deutschen Daten ebenfalls davon ausgegangen werden, dass ein höheres laufendes Einkommen zu einer Erhöhung der Wahrscheinlichkeit dafür führt, dass jemand Wohneigentum besitzt. Dies gilt umso mehr für jüngere Haushalte. Für ältere Haushalte ist der Effekt weniger bedeutsam.

Bei Hoderlein und White (2009) kann der Zusammenhang für 1999/2001 über alle Einkommensbereiche hinweg festgestellt werden. Dies ist bei den deutschen Daten nicht der Fall, muss jedoch auch im Zusammenhang mit den geringeren Beobachtungszahlen in den Randbereichen der Einkommensverteilung gesehen werden. Für die häufigen Einkommen (Bereich zwischen 9 und 11) haben die geschätzten Funk-

tionen in der Regel einen steigenden Verlauf und unterscheiden sich insbesondere für die jüngeren Haushalte signifikant von null.

Ähnlich wie bei Hoderlein und White (2009), deutet alles darauf hin, dass der positive Effekt in Deutschland vor allem bei den mittleren Einkommen ausgeprägter ist. Für die unteren Einkommensschichten lässt sich der positive Effekt kaum bestätigen, denn wenn der Effekt positiv wird, dann in der Regel erst etwa ab Werten um 10. Für die höheren Einkommen zeigt sich, dass sich ein abnehmender Verlauf der Kurve in der Regel erst ab Werten von 11 ergibt. Insgesamt scheint es so als ob in Deutschland zunächst ein gewisses Mindesteinkommen gegeben sein muss, bevor ein Zuwachs an laufendem Einkommen die Wahrscheinlichkeit, Wohneigentum zu besitzen, erhöhen kann. Außerdem scheint das laufende Einkommen auch in den Hoheinkommensbereichen noch von Bedeutung zu sein. Beides spricht für eine in Deutschland vergleichsweise besonders vorsichtige Kreditvergabe. Insgesamt liefern die Ergebnisse für das Panel 1999/2001 somit deutliche Evidenz für das Vorhandensein einer Liquiditätsbeschränkung der deutschen Haushalte bezüglich des Erwerbs von Wohneigentum. Insbesondere für jüngere Haushalte scheint das laufende Einkommen eine besonders wichtige Rolle zu spielen.

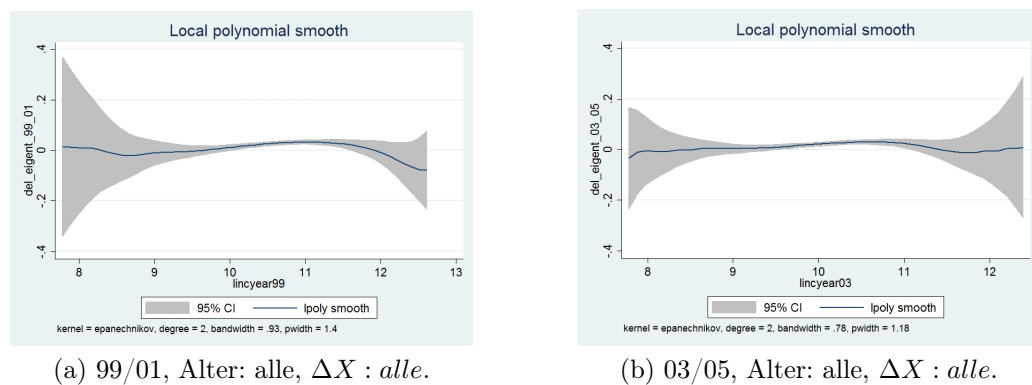


Abbildung 9: Ergebnisse für die Gruppe „alle“ 1999/2001 im Vergleich zu 2003/2005.
Quelle: Eigene Darstellung.

Fraglich ist nun, ob sich diese Ergebnisse ändern, wenn man das Panel 2003/2005 betrachtet.

Die Antwort lautet klar: Nein, nicht wesentlich, die Ergebnisse bleiben sehr ähnlich (Siehe Abbildung 9).⁶² Ein wesentlicher Unterschied ergibt sich lediglich für die als bereits wenig aussagekräftig identifizierte Altersgruppe „um den Durchschnitt“. Für

⁶² Vgl. Anhang A5 mit A6, bzw. betrachte Anhang A4.

die Einkommensschwankungen $+/- 0$ und $+/- 750$ kann nun ebenfalls ein signifikant von null verschiedener, positiver Effekt des Einkommens auf die marginale Wahrscheinlichkeit, Wohneigentum zu besitzen, gefunden werden. Zudem kann für die Altersgruppe „alt“ kein Effekt mehr für über sämtliche Einkommensänderungen hinweg gefunden werden.

Mit Ausnahme der Gruppe „um den Durchschnitt“ lassen sich somit insgesamt für das Panel 2003/2005 bezüglich der verschiedenen Altersgruppen die gleichen Aussagen treffen. Für jüngere Haushalte spielt das laufende Einkommen eine größere Rolle.

Was bedeutet dies nun für die Liquiditätsbeschränkung der Haushalte?

Um auf eine Lockerung der Liquiditätsbeschränkung deutscher Haushalte schließen zu können, hätten sich die gefundenen positiven Effekte des laufenden Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, deutlich abschwächen müssen. Bei Hoderlein und White (2009) ist dies der Fall: die ursprünglich positive marginale Wahrscheinlichkeit, Wohneigentum zu besitzen wird in 2003/2005 insignifikant. Die Liquiditätsbeschränkung verschwindet. Hiervon kann bei den deutschen Daten keine Rede sein. Weder ein Verschwinden der Liquiditätsbeschränkung, noch eine sich verringernde Liquiditätsbeschränkung kann gefunden werden: Ein allgemeines Absinken der marginalen Wahrscheinlichkeiten innerhalb des positiven Bereichs kann nicht festgestellt werden. Der Effekt des Einkommens auf die Wahrscheinlichkeit, ein Haus zu besitzen, bleibt positiv und signifikant von null verschieden, wenn er dies auch schon im vorangegangenen Panel war.⁶³

Für Deutschland kann somit geschlossen werden, dass es keine nennenswerten Änderungen bezüglich der Einstellung gegenüber dem laufenden Einkommen als Entscheidungsgrundlage für die Kreditvergabe gab. Die theoretischen Vorüberlegungen können bestätigt werden. Sowohl Verbraucher, als auch Kreditgeber räumen dem laufenden Einkommen nach wie vor eine wichtige Rolle ein. Einkommensdifferenzen und Einkommensrisiken bleiben entscheidend für die Vergabe von Hypotheken und den Erwerb von Wohneigentum.

⁶³ Vergleicht man die beiden Tabellen in Anhang A4, so ergibt sich nur einmal eine Änderung von „ja“ zu „nein“ und zwar bei der Gruppe „alt“.

4. Fazit

Im Theorieteil dieser Diplomarbeit wurde eine neue ökonometrische Methode von Hoderlein und White (2009) dargestellt. Diese ermöglicht den Umgang mit persistenter, unbeobachtbarer Heterogenität in allgemeinen nichtseparablen und nonparametrischen Konstellationen. Der Vorteil der Methode liegt in ihrer Flexibilität, den zahlreichen Erweiterungen und in der einfachen Anwendbarkeit im Binärmodell. Prinzipiell lässt sich der Ansatz überall dort nutzen, wo individuenspezifische Effekte, wie zum Beispiel Präferenzen, eine Rolle spielen.

Das Besondere an der Anwendung im Kontext des Wohneigentumserwerbs ist, dass nur eine sehr geringe Zahl von Variablen benötigt wird, da für sämtliche unbeobachtbare konstante Faktoren kontrolliert werden kann. Der marginale Effekt des laufenden Haushaltseinkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen, kann somit lediglich unter Zuhilfenahme einer einzigen Kontrollvariable, dem Alter des Haushaltsvorstands, geschätzt werden. Als ein Indikator für allgemeine Risiken im Bereich des Wohnimmobiliensektors bietet der Ansatz somit gewisse Vorteile gegenüber bisherigen Ansätzen. Außerdem setzt er nicht an den Immobilienpreisen, sondern auf Ebene der Haushalte an.

Während Hoderlein und White (2009) in ihrer Arbeit für die USA kurz vor der Immobilienkrise ein Verschwinden der Liquiditätsbeschränkung der Haushalte feststellen, liegen hierzu für die deutschen Daten keinerlei Hinweise vor. Es ergibt sich weder Evidenz für ein Verschwinden, noch für eine Reduktion der Liquiditätsbeschränkung deutscher Haushalte. Es ergibt sich ein in der Regel über die Zeit hinweg konstanter, positiver durchschnittlicher marginaler Effekt des Einkommens auf die Wahrscheinlichkeit, Wohneigentum zu besitzen. Im Gegensatz zu den USA, lässt sich dieser Effekt in Deutschland jedoch nicht über alle Einkommensbereiche hinweg, sondern nur im Bereich der mittleren Einkommen feststellen: Dort zeigt sich ein signifikant positiver marginaler Effekt mit steigendem Verlauf. Besonders ausgeprägt ist dieser für jüngere Haushalte.

Insgesamt spiegeln die Ergebnisse die in Deutschland im Betrachtungszeitraum im Vergleich zu den USA wesentlich vorsichtigeren Kreditvergabe wider. Das laufende Einkommen spielt eine durchweg wichtige Rolle. Von einer veränderten Risikoeinstellung kann keine Rede sein. Für Deutschland scheint der Ausdruck „No boom - no bust!“ zuzutreffen, so dass sich bisher noch kein gesamtwirtschaftliches Risiko aus dem Markt für Wohnimmobilien ergab.

Für die Zukunft wäre es jedoch interessant, den vorgestellten ökonometrischen Ansatz im Laufe der Zeit immer wieder neu anzuwenden und die gefundenen Effekte zu vergleichen. Etwaige Risiken könnten so unter Zuhilfenahme weiterer Indikatoren frühzeitig aufgedeckt werden. Denkbar wäre außerdem ein weiterführender, länderübergreifender Vergleich, denn die Anwendung des ökonometrischen Ansatzes für zusätzliche Länder erscheint aufgrund der wenigen benötigten Daten relativ unkompliziert. Unter Einbeziehung institutioneller Faktoren der jeweiligen nationalen Immobilienmärkte könnten weitere Einblicke in Bezug auf die Entstehung von Immobilienkrisen gewonnen werden. Als zu untersuchende Länder würden sich europäische Länder mit Immobilienkrisen, wie zum Beispiel Spanien, Großbritannien oder Irland, anbieten.

Literaturverzeichnis

Altonji, J. und R. Matzkin (2005). Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors. *Econometrica* 73: S. 1053-1103.

Arellano, M. (2003). Discrete Choice with Panel Data. *Investigaciones Economicas* 27: S. 423-458.

Arellano, M. und S. Bonhomme (2009). Identifying Distributional Characteristics in Random Coefficients Panel Data Models. CEMFI. Working Paper.

Arellano, M. und B. Honoré (2001). Panel Data Models: Some Recent Developments. In Heckman, J. und E. Leamer: *Handbook of Econometrics*, Vol. 5. North Holland.

Arellano, M. und J. Hahn (2007). Understanding Bias in Nonlinear Panel Models: Some recent developments. In: Blundell, R. W., W. K. Newey und P. Torsten: *Advances in Economics and Econometrics*, Vol. 3. Cambridge University Press.

Badunenko, O., N. Barasinska und D. Schäfer (2009). Genügend Widerstandskraft? Immobilienfinanzierung und Haushaltsverschuldung in Deutschland, Großbritannien, den Niederlanden und den USA. *Vierteljahreshefte zur Wirtschaftsforschung* 78(1): S. 127-142.

Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata*. Stata Press.

Belke, A. (2009). Die Auswirkungen der Geldmenge und des Kreditvolumens auf die Immobilienpreise: Ein ARDL-Ansatz für Deutschland. DIW Berlin. Discussion Paper 953.

Bester, A. und C. Hansen (2009). Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model. *Journal of Business and Economic Statistics* 27(2): S. 235-250.

Bierens, H. J. (2005). *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press.

Blundell, R. und J. Powell (2004). Endogeneity in Semiparametric Binary Response Models. *Review of Economic Studies* 71: S. 655-679.

Brown, S. und K. Taylor (2008). Household Debt and Financial Assets: Evidence from Germany, Great Britain and the USA. *Journal of the Royal Statistical Society* 171: S. 615-643.

Butler, J. S. und R. A. Moffit (1982). A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model. *Econometrica* 50: S. 761-764.

Cameron, A. C. und P. K. Trivedi (2005). *Microeconometrics - Methods and Applications*. Cambridge University Press.

Cameron, A. C. und P. K. Trivedi (2009). *Microeconometrics Using Stata*. Stata Press.

Chamberlain, G. (2010). Binary Response Models for Panel Data: Identification and Information. *Econometrica* 78: S. 159-168.

Chamberlain, G. (1984). Panel Data. In Griliches Z. und M. D. Intriligator. *Handbook of Econometrics*, Vol. 2. North Holland.

Chernozhukov, V., I. Fernandez-Val, J. Hahn und W. Newey (2009). Identification and Estimation of Marginal Effects in Nonlinear Panel Models. MIT. Working Paper.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 74. S. 829-836.

Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society Series B* 41: S. 1-31.

Debelle, G. (2004). Verschuldung der privaten Haushalte und gesamtwirtschaftliche Folgen. Bank für Internationalen Zahlungsausgleich. Quartalsbericht März 2004.

Elstrodt, J. (2005). *Maß- und Integrationstheorie*, 4. Auflage. Springer.

Fahrmeir, L., T. Kneib und S. Lang (2009). Regression - Modelle, Methoden und Anwendungen, 2. Auflage. Springer.

Fan, J. (1992). Design-adaptive Nonparametric Regression. Journal of the American Statistical Association 87: S. 998-1004.

Fan J. und I. Gijbels (1996). Local Polynomial Modelling and its Applications. Chapman and Hall.

Frick, J. R., S. P. Jenkins, D. R. Lillard, O. Lipps und M. Wooden (2008). Die internationale Einbettung des Sozio-ökonomischen Panels (SOEP) im Rahmen des Cross-National Equivalent File (CNEF). Vierteljahreshefte zur Wirtschaftsforschung 77(3): S. 110-129.

Gordon, G. (2008). The Subprime Panic. European Financial Management 15: S.10-46.

Girouard, N., M. Kennedy, P. van den Noord, C. Christophe (2006). Recent House Price Developments - The Role of Fundamentals. OECD Economics Department. Working Paper 475.

Gräf, B. und S. Schneider (2009). Wie bedrohlich sind die mittelfristigen Inflationsrisiken? Deutsche Bank Research. Aktuelle Themen 443.

Graham, B. und J. Powell (2008). Identification and Estimation of 'Irregular' Correlated Random Coefficient Models. NBER. Working Paper 14469.

Hahn, J. und W. Newey (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Data Models. Econometrica 72: S. 1295-1319.

Haisken-DeNew, J. P. und J. R. Frick (2005). DTC-Desktop Companion to the German Socio-Economic Panel (SOEP). Updated to wave 21(U), Version 8.0.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

- Härdle, W. und J. S. Marron (1985). Optimal Bandwidth Selection in Nonparametric Regression Function Estimation. *Annals of Statistics* 13: S. 1465-1481.
- Hausman, J., B. Hall und Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica* 52: S. 909-938.
- Headey, B. und E. Holst (2008). A Quarter Century of Change: Results from the German Socio-Economic Panel (SOEP). DIW Berlin. SOEP Wave Report 1-2008.
- Hoderlein, S. (2005). Nonparametric Demand Systems, Instrumental Variables and a Heterogeneous Population. Brown University. Working Paper.
- Hoderlein, S. und H. White (2009). Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects. *Cemmap*. Working Paper 33/09.
- Honoré, B. (1992). Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. *Econometrica* 60: S. 533-565.
- Honoré, B. und E. Kyriazidou (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica* 68: S. 839-874.
- Honoré, B. und E. Tamer (2006). Bounds on Parameters in Dynamic Discrete Choice Models without Strict Exogeneity. *Econometrica* 74: S. 611-629.
- Kholodilin, K. A., J.-O. Menz und B. Siliverstovs (2008). Immobilienkrise? Warum in Deutschland die Preise seit Jahrzehnten stagnieren. DIW Berlin Wochenbericht 17: S. 214-220.
- Kyriazidou, E. (1997). Estimation of a Panel Data Sample Selection Model. *Econometrica* 65: 1335-1364.
- Lancaster, T. (2000). The Incidental Parameter Problem since 1948. *Journal of Econometrics* 95: S.391-413.
- Lee, M.-J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. Springer-Verlag.

- MacLennan, D., Muellbauer, J. und Stephens, M. (1998). Asymmetries in Housing and Financial Market Institutions and EMU. *Oxford Economic Policy* 14(3): S. 54-80.
- Manski, C. F. (1975). Maximum Score Estimation of the Stochastic Utility Model of Choice. *Journal of Econometrics* 3: S. 205-228.
- Manski, C. F. (1987). Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data. *Econometrica* 55(2): S. 357-362.
- Myoung-jae, L. (2006). *Method of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, 1. Auflage. Springer.
- Myoung-jae, L. (2010). *Microeconometrics - Method of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, 2. Auflage. Springer.
- Newey, W. (1994). Kernel Estimation of Partial Means and a General Variance Estimator. *Econometric Theory* 10: S. 233-253.
- Rasch, G. (1961). On the General Law and the Meaning of Measurement in Psychology. *Proceedings of the Fourth Berkely Symposium on Mathematical Statistics and Probability*, Vol. 4. Berkeley UC Press.
- Rasch, G. (1960). Probabilistic Models for some Intelligence and Attainment Tests. *Denmarks Paedagogiske Institut* Kopenhagen.
- Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics* 27: S. 832-837.
- Ruppert, D. und M. P. Wand (1994). Multivariate Locally Weighted Least Squares Regression. *Annals of Statistics* 22: S. 1346-1370.
- Sinn, H.-W. (2010). *Kasino Kapitalismus - Wie es zur Finanzkrise kam, und was jetzt zu tun ist*. Ullstein.

Statistisches Bundesamt (2010). Statistisches Jahrbuch 2010 - Für die Bundesrepublik Deutschland mit „Internationalen Übersichten“. Druck-Buch-Verlag.

Wagner, G. C., J. Göbel, P. Krause und R. Pischner (2008). Das Sozio-ökonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland: Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender). AStA Wirtschafts- und Sozialstatistisches Archiv 2: S. 301-328.

Weber, A. A. (2009). Regulatorische und institutionelle Konsequenzen aus der Finanzkrise. Gastvortrag in der Vortragsreihe der Stiftung Kreditwirtschaft an der Universität Hohenheim am 11. November 2009.

Wooldridge, J. M. (2002). Econometric Analysis of Cross section and Panel Data. MIT-Press.

Wooldridge, J. M. (2006). Introductory Econometrics: A Modern Approach, 3rd Edition. Thomson/South-Western.

A. Anhang

A.1. Deskriptive Statistiken für die Originaldaten

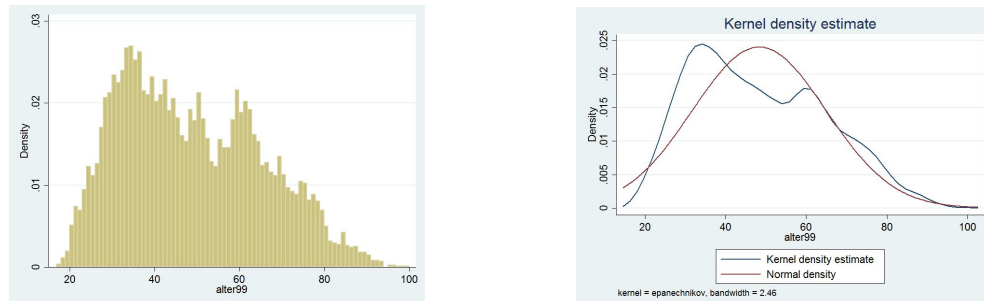


Abbildung A1: Verteilung Alter 99.
Quelle: Eigene Darstellung.

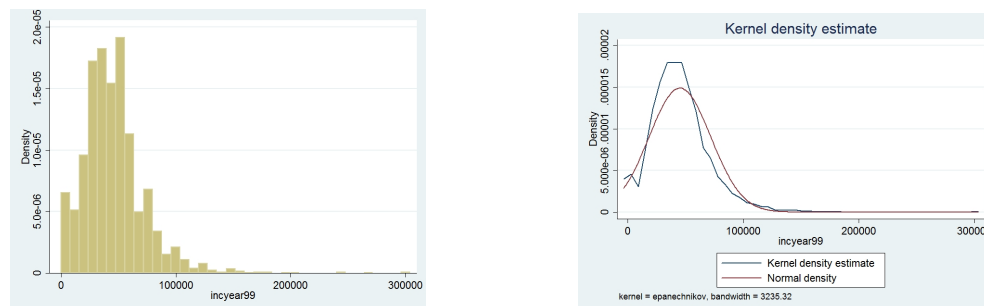


Abbildung A2: Verteilung der Haushalts-Nettoeinkommen 1999.
Quelle: Eigene Darstellung.

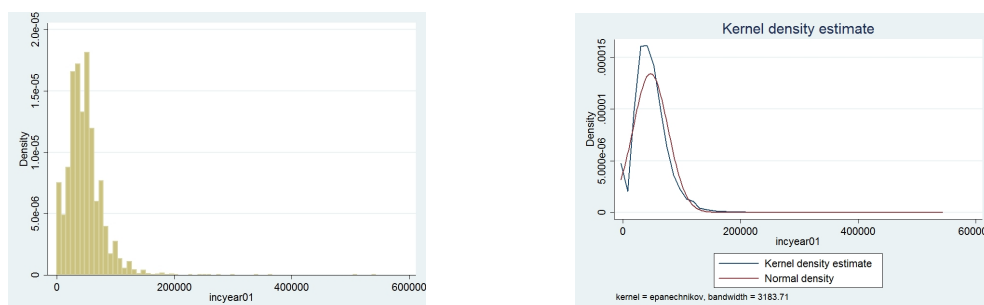


Abbildung A3: Verteilung der Haushalts-Nettoeinkommen 2001.
Quelle: Eigene Darstellung.

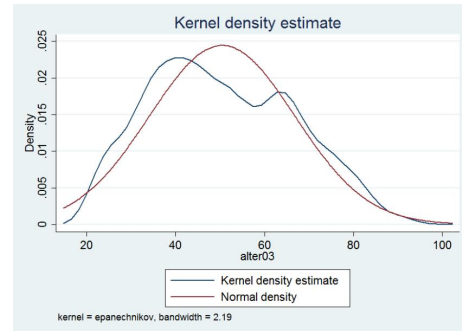
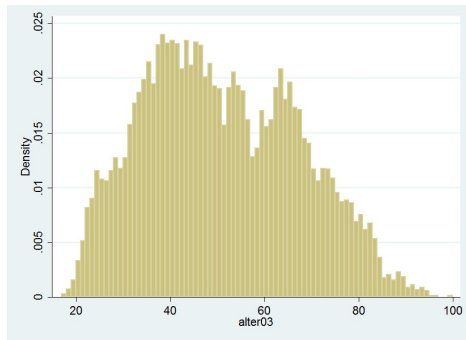


Abbildung A4: Verteilung Alter 2003.
Quelle: Eigene Darstellung.

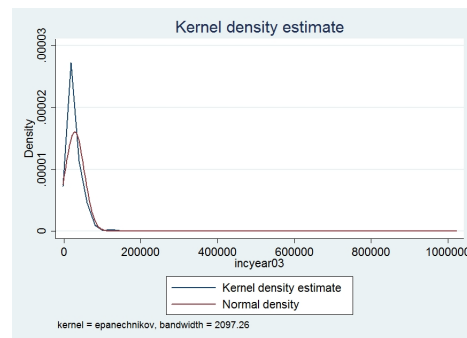
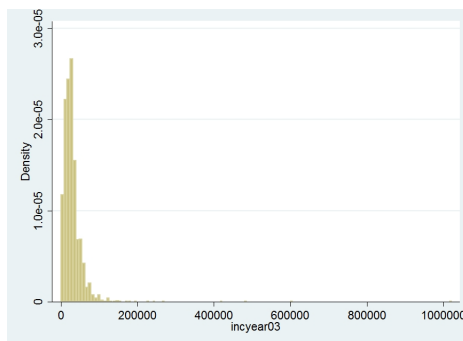


Abbildung A5: Verteilung der Haushalts-Nettoeinkommen 2003.
Quelle: Eigene Darstellung.

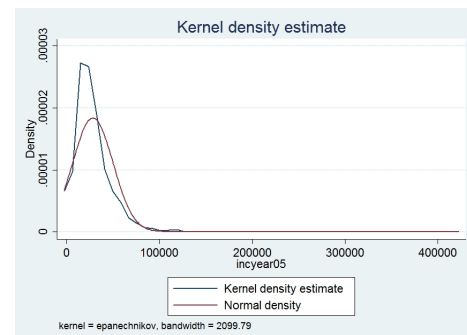
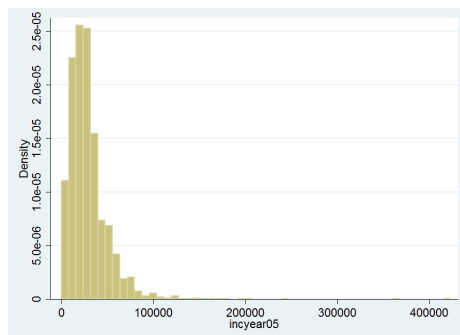


Abbildung A6: Verteilung der Haushalts-Nettoeinkommen 2005.
Quelle: Eigene Darstellung.

A.2. Deskriptive Statistiken für die aufbereiteten Daten

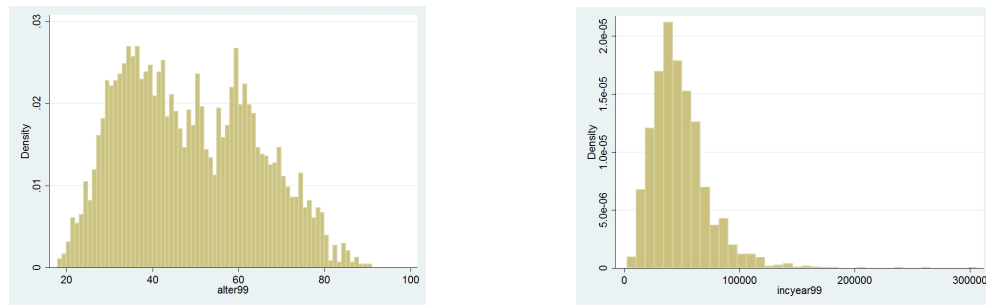


Abbildung A7: Verteilung Alter 1999 (links),
Verteilung der Haushalts-Nettoeinkommen 1999 (rechts).
Quelle: Eigene Darstellung.

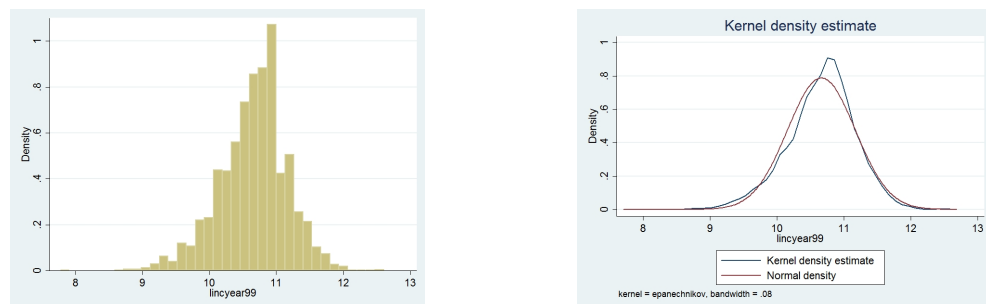


Abbildung A8: Verteilung der logarithmierten
Haushalts-Nettoeinkommen 1999.
Quelle: Eigene Darstellung.

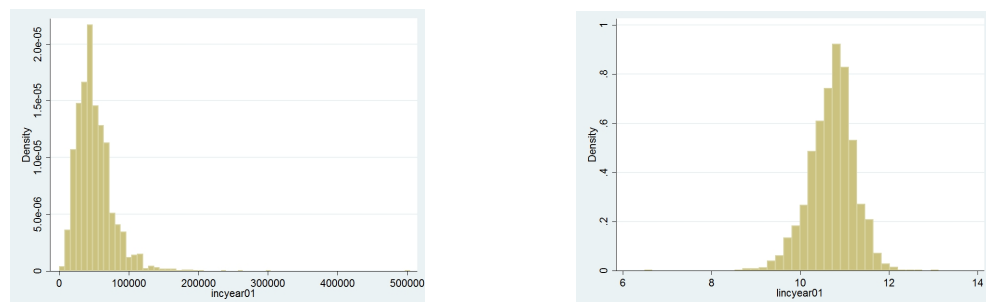


Abbildung A9: Verteilung der Haushalts-Nettoeinkommen 2001 (links),
Verteilung der logarithmierten Haushalts-Nettoeinkommen 2001 (rechts).
Quelle: Eigene Darstellung.

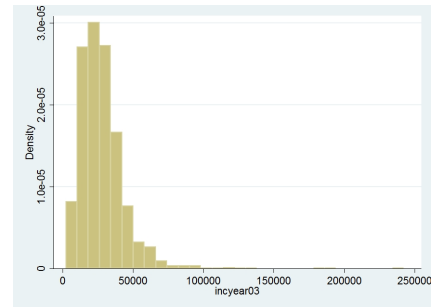
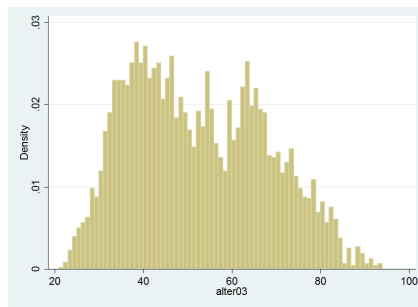


Abbildung A10: Verteilung Alter 2003 (links),
Verteilung Haushalts-Nettoeinkommen 2003 (rechts).
Quelle: Eigene Darstellung.

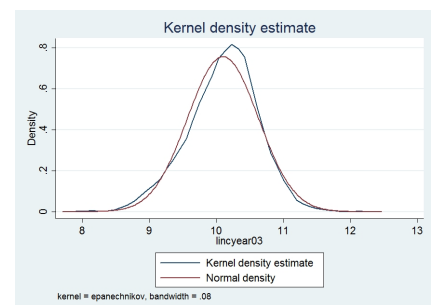
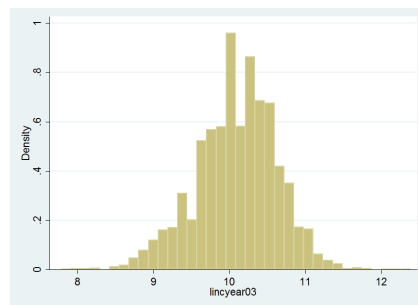


Abbildung A11: Verteilung der logarithmierten
Haushalts-Nettoeinkommen 2003.
Quelle: Eigene Darstellung.

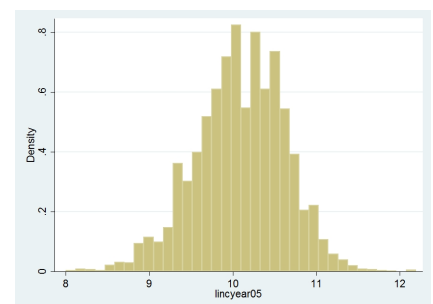
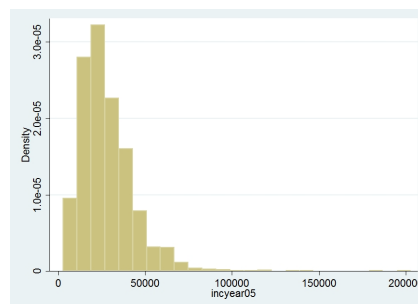


Abbildung A12: Verteilung der Haushalts-Nettoeinkommen 2005 (links),
Verteilung der logarithmierten Haushalts-Nettoeinkommen 2005 (rechts).
Quelle: Eigene Darstellung.

A.3. Resultate für die USA

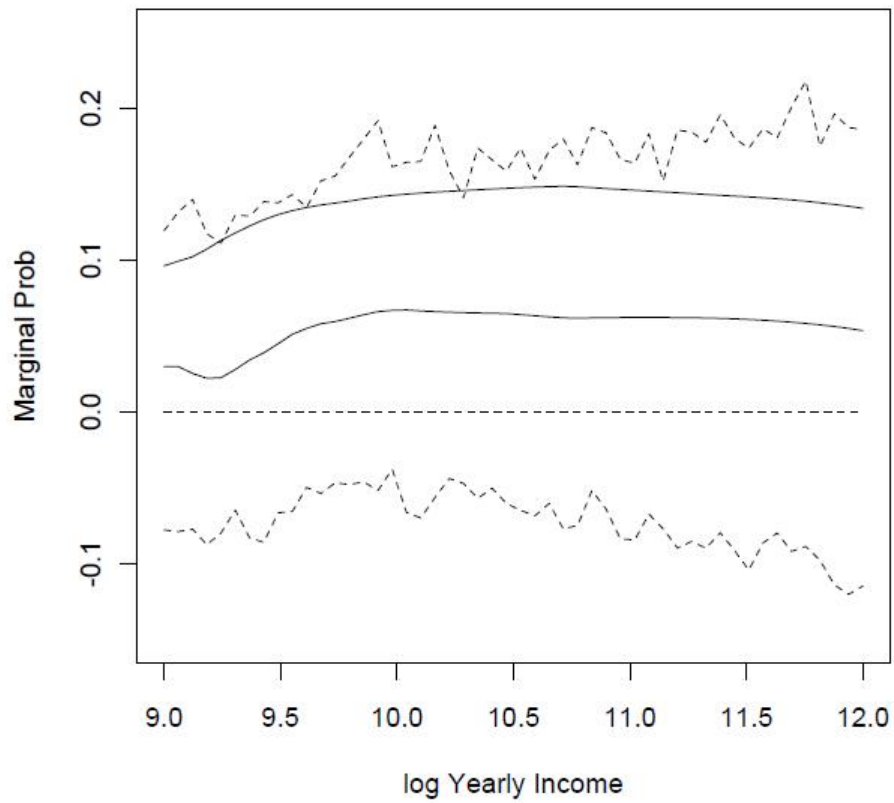


Abbildung A13: Marginale Wahrscheinlichkeit für den Besitz von Wohneigentum bei gegebenem Einkommen (Zeitraum: 1999/2001 und 2003/2005).

Quelle: Hoderlein und White (2009).

A.4. Übersicht Ergebnisse Deutschland

Tabelle A1: Ergebnisübersicht Panel 1999/2001

Altersgruppe: alle	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	401	ja
$\Delta X = +/ - 750$	575	ja
$\Delta X = +/ - 3.000$	1.493	ja
$\Delta X = \text{alle}$	4.787	ja
Altersgruppe: „jung“ (25-40)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	106	ja
$\Delta X = +/ - 750$	143	ja
$\Delta X = +/ - 3.000$	376	ja
$\Delta X = \text{alle}$	1.790	ja
Altersgruppe: „alt“ (50-70)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	154	nein
$\Delta X = +/ - 750$	227	nein
$\Delta X = +/ - 3.000$	606	nein
$\Delta X = \text{alle}$	1.711	ja
Altersgruppe: „um Durchschnitt“ (41-55)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	109	nein
$\Delta X = +/ - 750$	132	nein
$\Delta X = +/ - 3.000$	349	nein
$\Delta X = \text{alle}$	1.328	ja

Anmerkung: Mit „Effekt signifikant“ ist hier gemeint:
Die Funktion ist bei einem Signifikanzniveau von 5% an mindestens einem Punkt innerhalb des betrachteten Einkommensspektrums von null verschieden.

Quelle: Eigene Darstellung.

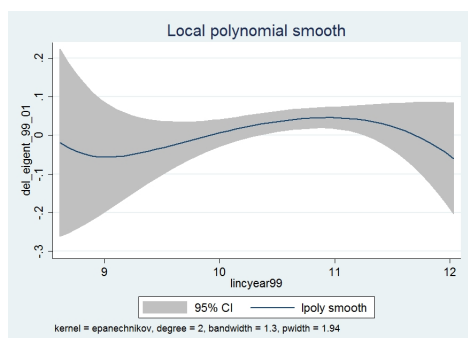
Tabelle A2: Ergebnisübersicht Panel 2003/2005

Altersgruppe: alle	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	457	ja
$\Delta X = +/ - 750$	1.065	ja
$\Delta X = +/ - 3.000$	2.415	ja
$\Delta X = \text{alle}$	4.787	ja
Altersgruppe: „jung“ (25-40)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	105	ja
$\Delta X = +/ - 750$	180	ja
$\Delta X = +/ - 3.000$	548	ja
$\Delta X = \text{alle}$	1.337	ja
Altersgruppe: „alt“ (50-70)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	210	nein
$\Delta X = +/ - 750$	468	nein
$\Delta X = +/ - 3.000$	976	nein
$\Delta X = \text{alle}$	1.795	nein
Altersgruppe: „um Durchschnitt“ (41-55)	Zahl der Haushalte	Effekt signifikant?
$\Delta X = +/ - 0$	124	ja
$\Delta X = +/ - 750$	226	ja
$\Delta X = +/ - 3.000$	637	nein
$\Delta X = \text{alle}$	1.496	ja

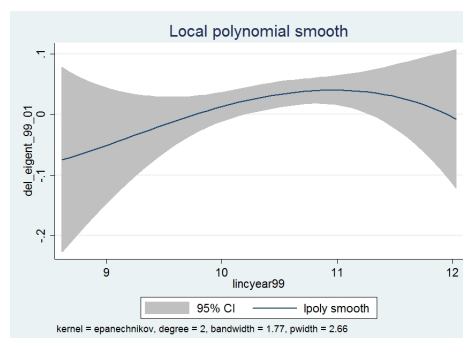
Anmerkung: Mit „Effekt signifikant“ ist hier gemeint:
 Die Funktion ist bei einem Signifikanzniveau von 5% an mindestens einem Punkt
 innerhalb des betrachteten Einkommensspektrums von null verschieden.

Quelle: Eigene Darstellung.

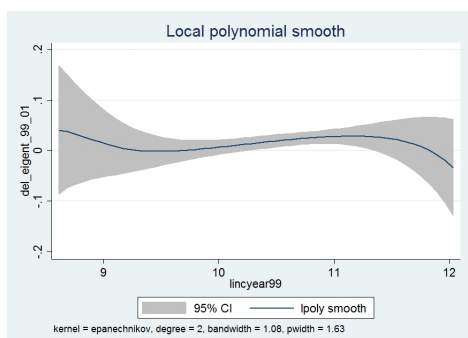
A.5. Ergebnisse für Deutschland 1999/2001 im Detail



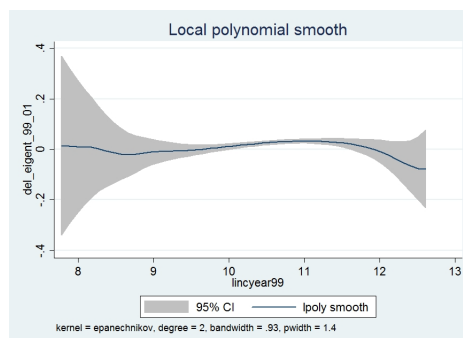
(a) 99/01, Alter: alle, $\Delta X : +/ - 0$.



(b) 99/01, Alter: alle, $\Delta X : +/ - 750$.



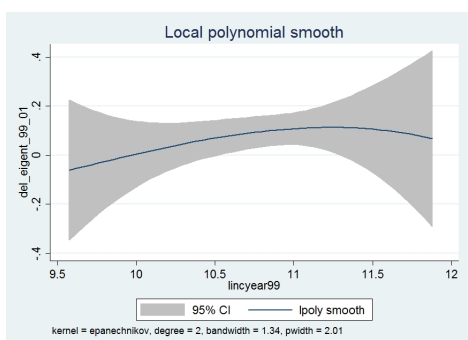
(c) 99/01, Alter: alle, $\Delta X : +/ - 3.000$.



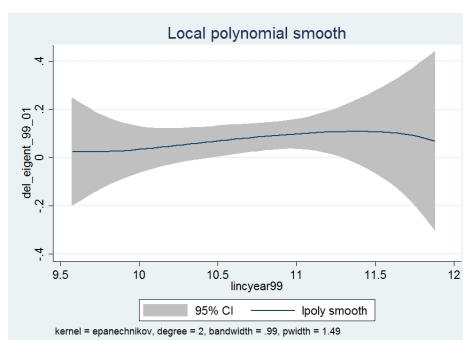
(d) 99/01, Alter: alle, $\Delta X : alle$.

Abbildung A14: Ergebnisse 99/01, alle Altersgruppen.

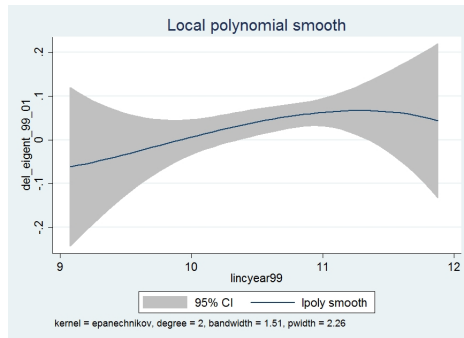
Quelle: Eigene Darstellung.



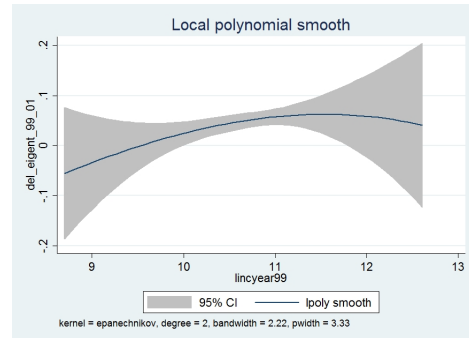
(a) 99/01, Alter: 25-40, $\Delta X : +/ - 0$.



(b) 99/01, Alter: 25-40, $\Delta X : +/ - 750$.

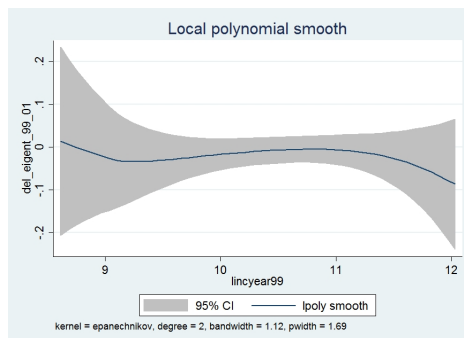


(c) 99/01, Alter: 25-40, ΔX : ± 3.000 .

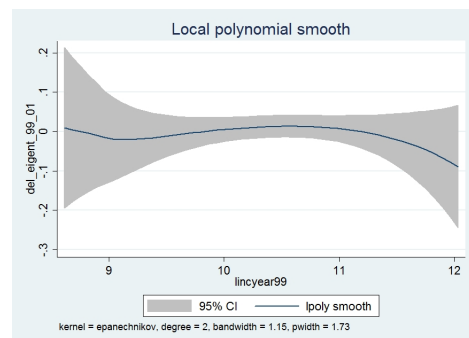


(d) 99/01, Alter: 25-40, ΔX : *alle*.

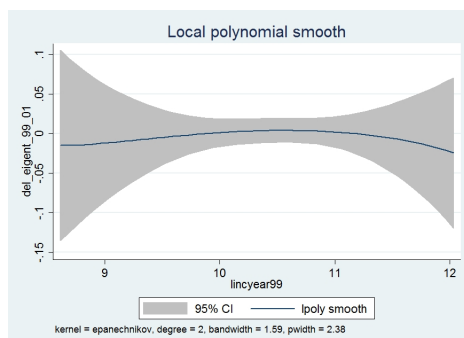
Abbildung A15: Ergebnisse 99/01, Altersgruppe „jung“.
Quelle: Eigene Darstellung.



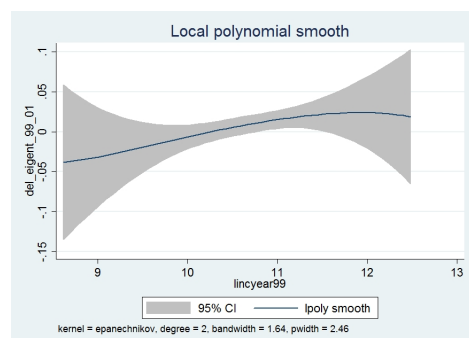
(a) 99/01, Alter: 50-70, ΔX : ± 0 .



(b) 99/01, Alter: 50-70, ΔX : ± 750 .

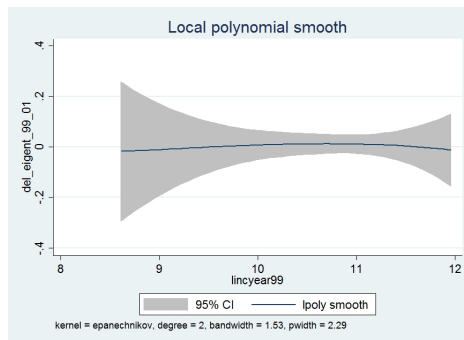


(c) 99/01, Alter: 50-70, ΔX : ± 3.000 .

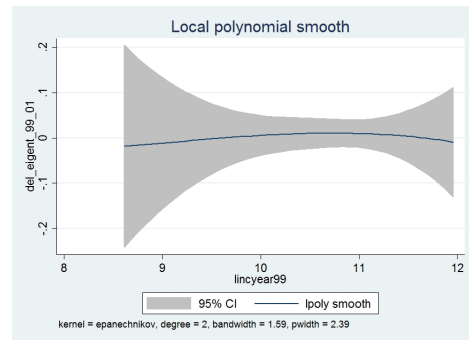


(d) 99/01, Alter: 50-70, ΔX : *alle*.

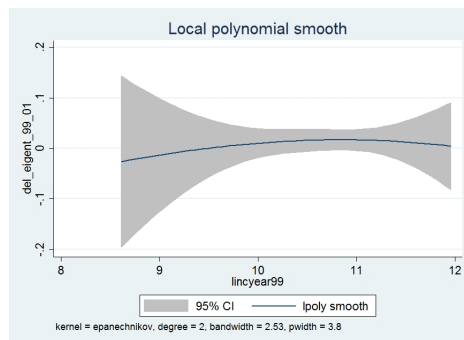
Abbildung A16: Ergebnisse 99/01, Altersgruppe „alt“.
Quelle: Eigene Darstellung.



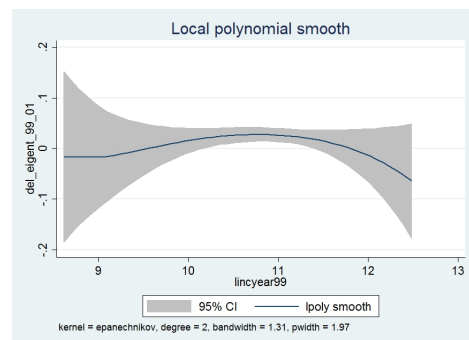
(a) 99/01, Alter: 41-55, $\Delta X : + / - 0$.



(b) 99/01, Alter: 41-55, $\Delta X : + / - 750$.



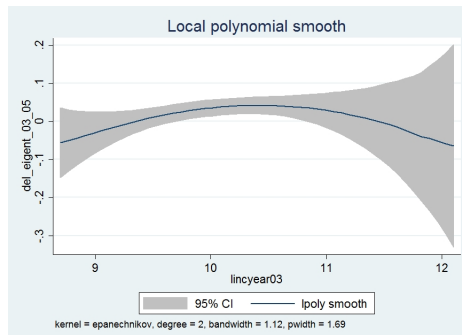
(c) 99/01, Alter: 41-55, $\Delta X : + / - 3.000$.



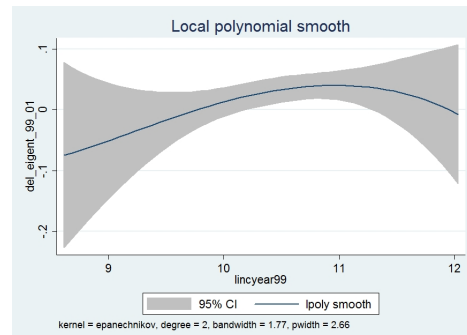
(d) 99/01, Alter: 41-55, $\Delta X : alle$.

Abbildung A17: Ergebnisse 99/01, Altersgruppe „um Durchschnitt“.
Quelle: Eigene Darstellung.

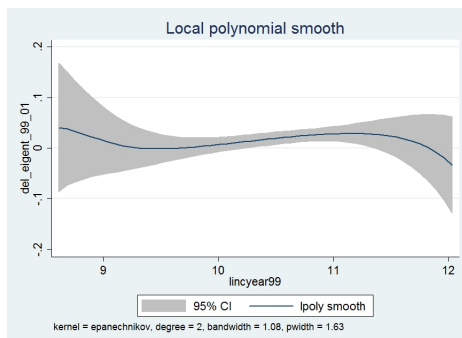
A6: Ergebnisse für Deutschland 2003/2005 im Detail



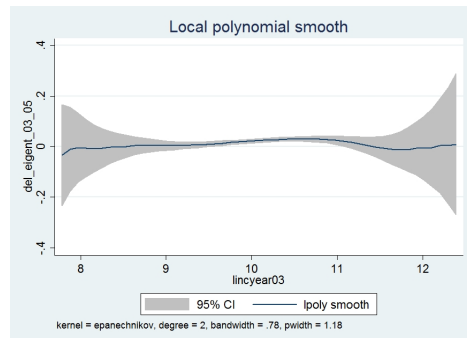
(a) 03/05, Alter: alle, $\Delta X : +/- 0$.



(b) 03/05, Alter: alle, $\Delta X : +/- 750$.



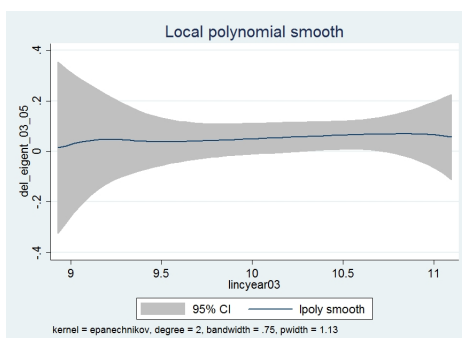
(c) 03/05, Alter: alle, $\Delta X : +/- 3.000$.



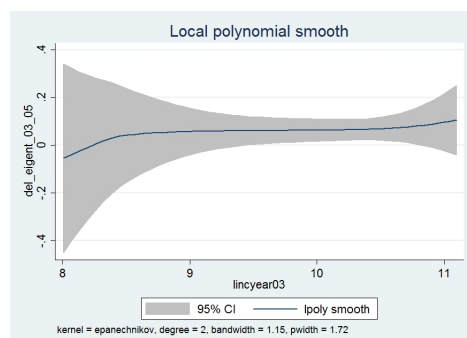
(d) 03/05, Alter: alle, $\Delta X : alle$.

Abbildung A18: Ergebnisse 03/05, alle Altersgruppen.

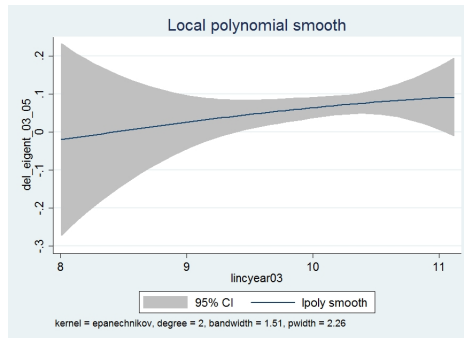
Quelle: Eigene Darstellung.



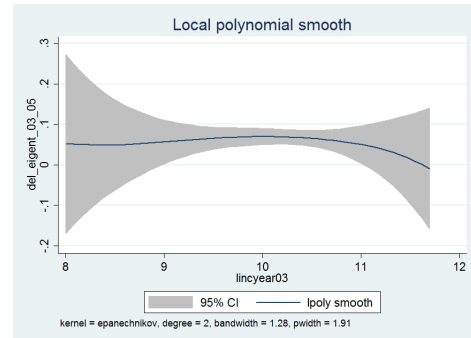
(a) 03/05, Alter: 25-40, $\Delta X : +/- 0$.



(b) 03/05, Alter: 25-40, $\Delta X : +/- 750$.

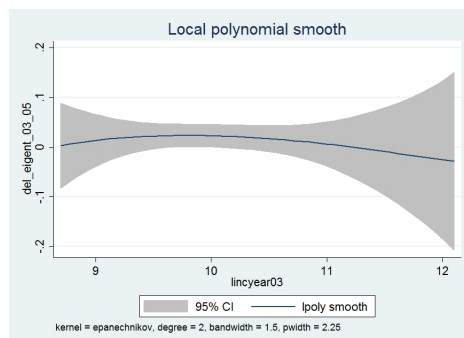


(c) 03/05, Alter: 25-40, ΔX : ± 3.000 .

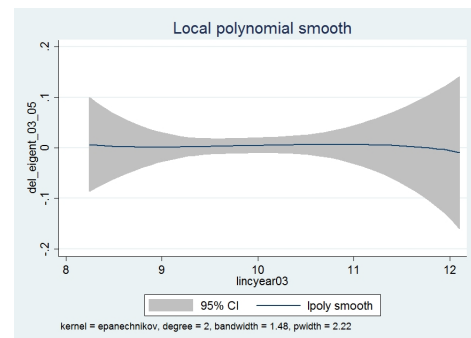


(d) 03/05, Alter: 25-40, ΔX : *alle*.

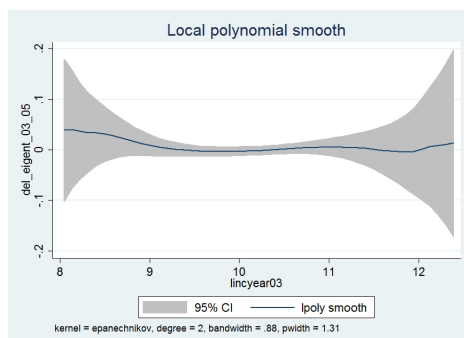
Abbildung A19: Ergebnisse 03/05, Altersgruppe „jung“.
Quelle: Eigene Darstellung.



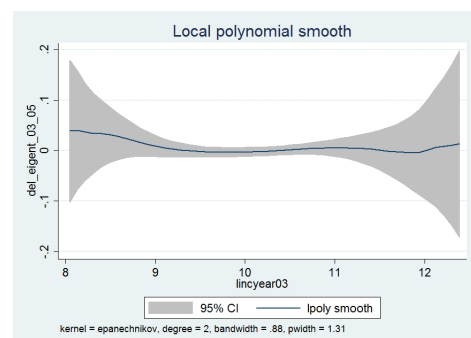
(a) 03/05, Alter: 50-70, ΔX : ± 0 .



(b) 03/05, Alter: 50-70, ΔX : ± 750 .

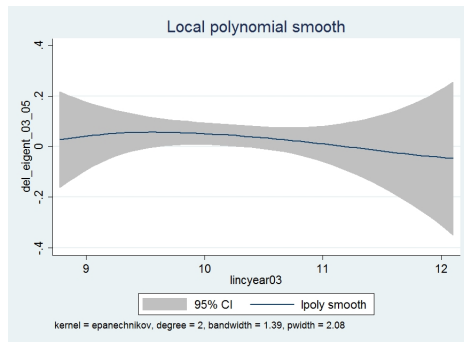


(c) 03/05, Alter: 50-70, ΔX : ± 3.000 .

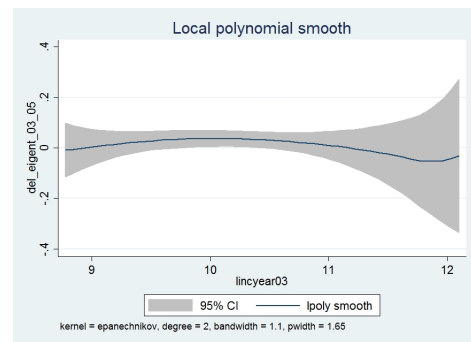


(d) 03/05, Alter: 50-70, ΔX : *alle*.

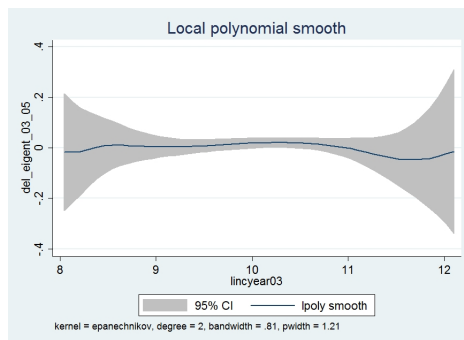
Abbildung A20: Ergebnisse 03/05, Altersgruppe „alt“.
Quelle: Eigene Darstellung.



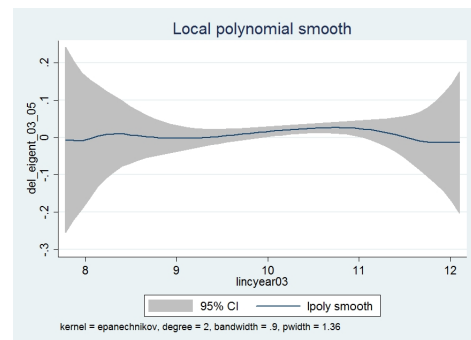
(a) 03/05, Alter: 41-55, $\Delta X : + / - 0$.



(b) 03/05, Alter: 41-55, $\Delta X : + / - 750$.



(c) 03/05, Alter: 41-55, $\Delta X : + / - 3.000$.



(d) 03/05, Alter: 41-55, $\Delta X : alle$.

Abbildung A21: Ergebnisse 03/05, Altersgruppe „um Durchschnitt“.

Quelle: Eigene Darstellung.